Web-based Multi-Party Computation
with Application to Anonymous Aggregate Compensation Analytics

Azer Bestavros

Computer Science Department
Hariri Institute for Computing
Boston University

Dataverse Privacy Workshop @ Harvard University
July 13, 2016

MPC in the land of social science with mayors, lawyers, CTOs, CIOs, administrators, politicians, journalists, and lawmakers...

A True Story by Azer Bestavros
October 31, 2013

MPC in the Real World: Dataverse Privacy Workshop @ Harvard University, July 13, 2016


December 11, 2013

100% TALENT
The Boston Women’s Compact

To make Greater Boston the premier place to work for women in America, by closing the wage gap and removing the visible and invisible barriers to women’s advancement. By doing so, we will build a more equitable workplace where all talent is hired, retained and valued.

GOAL 3
Evaluating Success

Employers agree to participate in a biennial review to discuss successes and challenges, as well as 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. The specific data to be reported will build on data already required by federal and state authorities and should not create an additional reporting burden.
September 4, 2014

Snapshot of (a small subset of) BWWC meetings from Azer’s Exchange Calendar

<table>
<thead>
<tr>
<th>Subject</th>
<th>Start</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cathy Minihan</td>
<td>Fri 9/5/2014 10:30 AM</td>
<td>2 hours</td>
</tr>
<tr>
<td>Simmons College</td>
<td>Mon 10/27/2014 10:30 AM</td>
<td>1.5 hours</td>
</tr>
<tr>
<td>Data Collection for Pay Equity</td>
<td>Tue 12/2/2014 11:30 AM</td>
<td>30 minutes</td>
</tr>
<tr>
<td>Simmons College people</td>
<td>Fri 1/24/2015 1:00 PM</td>
<td>30 minutes</td>
</tr>
<tr>
<td>Invitation: 100% Talent Discussion with Data Partners @ Tue Mar 17, 2015 2pm - 3pm (<a href="mailto:johns1k@simmons.edu">johns1k@simmons.edu</a>)</td>
<td>Tue 3/17/2015 2:00 PM</td>
<td>1 hour</td>
</tr>
<tr>
<td>Updated Invitation: MassMutual call with Harris Institute re: Data Collection, @ Thu May 14, 2015... Thu 5/14/2015 3:00 PM</td>
<td>Thu 5/14/2015 2:30 PM</td>
<td>1 hour</td>
</tr>
<tr>
<td>Invitation: Mock collection #1 @ Tue May 19, 2015 11am - 12pm (<a href="mailto:johns1k@simmons.edu">johns1k@simmons.edu</a>)</td>
<td>Tue 5/19/2015 11:00 AM</td>
<td>1 hour</td>
</tr>
<tr>
<td>Invitation: Mock Collection #2 @ Tue May 26, 2015 11am - 12pm (<a href="mailto:johns1k@simmons.edu">johns1k@simmons.edu</a>)</td>
<td>Tue 5/26/2015 11:00 AM</td>
<td>1 hour</td>
</tr>
<tr>
<td>Invitation: Mock Collection #3 @ Thu May 28, 2015 11am - 12pm (<a href="mailto:johns1k@simmons.edu">johns1k@simmons.edu</a>)</td>
<td>Thu 5/28/2015 11:00 AM</td>
<td>1 hour</td>
</tr>
<tr>
<td>Invitation: Call with BWWC @ Wed Jun 3, 2015 11:30AM - 12pm (<a href="mailto:johns1k@simmons.edu">johns1k@simmons.edu</a>)</td>
<td>Wed 6/3/2015 11:30 AM</td>
<td>30 minutes</td>
</tr>
<tr>
<td>Updated Invitation: 100% Talent Data Collection: Harris and Raytheon, @ Fri Jun 5, 2015 9am - 10am (<a href="mailto:johns1k@simmons.edu">johns1k@simmons.edu</a>)</td>
<td>Fri 6/5/2015 9:00 AM</td>
<td>1 hour</td>
</tr>
<tr>
<td>Invitation: 100% TALENT DATA COLLECTION: Harris &amp; Raytheon, @ Mon Jun 8, 2015 9am - 10:30am (<a href="mailto:johns1k@simmons.edu">johns1k@simmons.edu</a>)</td>
<td>Mon 6/8/2015 9:00 AM</td>
<td>1.5 hours</td>
</tr>
<tr>
<td>Invitation: Meeting with Boston Women’s Workforce Council @ Tue Aug 11, 2015 10am - 11am (<a href="mailto:johns1k@simmons.edu">johns1k@simmons.edu</a>)</td>
<td>Tue 8/11/2015 10:00 AM</td>
<td>1 hour</td>
</tr>
</tbody>
</table>

April 14, 2015

[Image]
June 8, 2015
D-day

“Katie, if this does not work out, I will just fax you the spreadsheet for you to enter...”
The congresswoman, who had signed onto a bill addressing income disparity between men and women, was impressed by the relevance he outlined. “It’s linking it back for the members of Congress,” Clark said. “Nobody would think, oh, the Paycheck Fairness Act, how is that tied into NSF funding?”

The meeting was slated for 15 minutes. It lasted 25.
April 28, 2016

The Sequel

- Compact doubled in size
- More elaborate analytics
- Hardened user interface
- Provide local sanity checks
- Provide comparative metrics

Multi-Party Computation (MPC)

- What is it?
  - Given multiple parties $p_1, p_2, \ldots, p_n$, each with private data $x_1, x_2, \ldots, x_n$
  - Engage parties in a protocol to compute a function $f(x_1, x_2, \ldots, x_n)$
  - without revealing more than the outputs of $f$

- State of the Art
  - Theory known for many decades, starting with Shamir’s “How to share a secret” in 1979
  - Frameworks and libraries increasingly available over the last few years, e.g., VIFF, Sharemind, Oblivm, ...
  - Experience with use cases involving real applications is limited and deployments are not easily portable
Example: Sum of Secret Records

Players split secrets into “shares”
All players exchange shares

Each player the sum to obtain a share of the sum of secrets!
Exchange/combine to get result

Lo and behold

Sum = 160  Sum = 160
The Parties in our MPC Setting

**Contributors**
- Have private data needed for computing the analytic
- Number of contributors is unknown in advance

**Analyzers**
- Ultimate recipient of the output of the analytic
- May also help in computing the analytic

**Service Provider**
- Connects/coordinate largely decoupled parties
- Has capacity to (partially) compute the analytic

Trust Assumptions

**Contributors & analyzers place some trust in each other**
- Analyzers trust that contributors will submit valid data
- Contributors trust that analyzers will protect aggregate output
- Contributors trust that analyzers will not collude with others

**No trust in service provider**
- Service provider cannot be entrusted with data or with the results of the computation
- Assume that service provider is incentivized to perform the computation on behalf of the contributors and analyzers
Operational Constraints

**Comprehensibility of the Protocol**
- MPC protocol must be simple for users (and lawyers/executives) to understand (and approve)

**Auditability of the Service**
- All software and processes must be transparent, with open-source code for outside auditing

**Capability of Contributor/Analyzers Infrastructures**
- Software clients must require no setup, no specialized software/hardware, or even public IP

**Simplicity of Client Interfaces**
- Usable by employees only familiar with spreadsheet application and web browsers

**Asynchronicity**
- Contributors need to be online while entering their data; analyzers to start/finish process

**Idempotence**
- Contributors must be able to resubmit/update their data to recover from errors, crashes, etc.

**Feedback**
- Interface for contributors must provide means to alert human users about spurious data

---

Putting it Together: Our Protocol

Let $G$ be an appropriate additive group such as $\mathbb{Z}/2^{64}\mathbb{Z}$ and distinguish each contributor using an index $i \in \{1, \ldots, n\}$. We call a single execution of the protocol a "session" and it proceeds in the following way:

1. the analyzer initiates the process by generating a secret and public RSA key pair $(s, p)$ and a unique session identifier $j \in \mathbb{N}$, submitting $p$ to the service provider, and sending $j$ to all the contributors;
2. each of the $n$ contributors possesses a secret data value $d_i \in G$ and does the following at least once:
   a. generate a secret random mask $m_i \in G$ and calculate the masked data $r_i = d_i + m_i$.
   b. send $r_i$ to the service provider and retrieve $p$ from the service provider to send back $c_i = \text{Enc}_p(m_i)$;
3. the service provider computes the sum of the masked data values to obtain the aggregate masked data quantity $R = \sum_{i=1}^{n} r_i$;
4. the analyzer then retrieves $R$ and all the $c_1, \ldots, c_n$ from the service provider, computes $m_i = \text{Dec}_s(c_i)$ for all $i$, computes $M = \sum_{i=1}^{n} m_i$, and obtains the final result $R - M = \sum_{i=1}^{n} d_i$. 
Web-Based MPC Platform

Bigger Picture: MPC as a Service

• Observations:
  – Data analysts should not be expected to worry about privacy or performance tradeoffs
  – Unrealistic to expect cryptographers to rewrite the huge existing corpus of privacy-agnostic analytics
  – Need to provide separate levers to manage privacy, utility, and performance tradeoffs

• Our Hypothesis (and research agenda):
  – MPC must be integrated seamlessly into popular, scalable data analytics platforms, e.g., MapReduce and Musketeer
MPC is great! But...

- MPC = math turtle shell for confidential data!
- Practical MPC frameworks exist, but
  - They are slow (not unlike turtles)
  - Learning curve is steep (trust me!)
- Lots of special-purpose gadgetery (circuits)
  - Apply to very narrow functions
  - Not for human consumption
MapReduce (MR): fast, like a hare!

- A programming paradigm for data analytics
  - Very easy to us (I teach it one lecture)
- Supported by distributed & elastic backend
  - Very easy to deploy (thanks to the cloud)
- Performance is ridiculously fast
  - 200 node cluster sorted 100TB of data in 23 minutes; think what 8,000 nodes can do!

Our Research

Separation of concerns
  - Domain experts specify analytics as MR code
  - Lawyers specify confidentiality constraints on data

Automate the interaction
  - Compile the MR code subject to constraints
  - Expose cost-privacy tradeoffs to resolve tussle

Build an execution platform
  - Extend MR backend (SPARC) to support MPC
  - Act as a platform to plug in special MPC gadgetry
The Scather Platform

A Programming Language
– To specify MR and MPC operations
– Not meant for consumption by non-experts

A Compiler
– To convert MR programs into Scather programs
– To expose tradeoffs

Backend Platform
– To run the show

Data Secured in Local Safes
+ MapReduce Local Infrastructure

+ Local MPC Client
+ Local Agent to Manage MR-MPC

+ Controller for MP Coordination
MPC as a Service

MPC in the Real World

- Boston Globe: Mayor Walsh pushes to gather data on wage gap
  https://goo.gl/B7Kj79
- BU Today: Computational Thinking Breaks a Logjam
  http://goo.gl/dnsqbo
- BU Research Magazine: Calculating Gender Pay Equity
  http://goo.gl/y6hIWH
- NPR OnPoint: Will Data Help Close The Gender Pay Gap?
  http://goo.gl/2J1thb
- Boston Globe: More Boston businesses join drive to end gender wage gap
  https://goo.gl/xbEKuX