Sharing Knowledge without Sharing Data

Platforms for resolving the false dichotomy between privacy and utility of information

Azer Bestavros

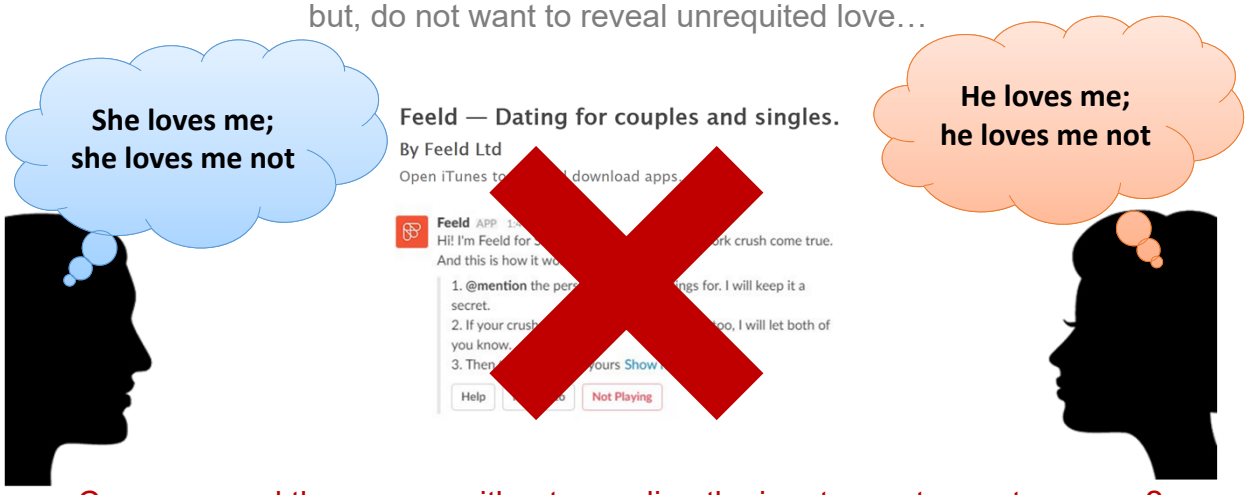
Computer Science Department
Hariri Institute for Computing
Boston University

RISE SICS Distributed Computing & Analytics Workshop

Electrum Kista, Stockholm, Sweden
September 26, 2018

The Valentine Question

Want to know if both parties are interested in each other
but, do not want to reveal unrequited love…



Can we reveal the answer without revealing the inputs – not even to an app?

(Yao's) Millionaires' Problem

Want to know who is wealthier

Can we reveal the answer without revealing the inputs – not even to an app?

The Dorm Access Question

Want to know if a student is allowed to access a dorm building

The New York Times SUBSCRIBE NOW LOG IN

N.Y. / REGION

On Campus, a Security Card and More

By AVI SALZMAN OCT. 5, 2003

KEYS have become passé at Quinnipiac University. The private school in Hamden went digital this year, installing hotel-style security locks on doors and allowing students to access everything from dorm rooms to meals using a card-swipe system that records their every move.

"It kinds of bothers me that the university can find out where students go and how long they stay by interrogating locks."

Can we let students in without knowing who they are?

The Labor Department Question

Want to know if companies like Google/Oracle are paying white men more


Department Of Labor

DOL HOME / NEWSROOM / NEWS RELEASES AND BRIEFS

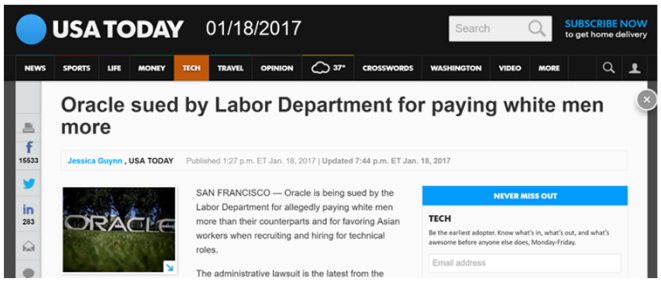
News Release
01/04/2017

Please note: As of January 20, 2017, information in some news releases may be out of date or not reflect current policies.

US DEPARTMENT OF LABOR SUES GOOGLE INC. FOR COMPENSATION DATA

Data requested as part of routine audit of a federal contractor

SAN FRANCISCO – The U.S. Department of Labor has filed a lawsuit to require Google Inc. to provide requested compensation data and documents for the multinational company's Mountain View headquarters as part of a routine compliance evaluation.


USA TODAY 01/18/2017

Oracle sued by Labor Department for paying white men more

Jessica Guynn, USA TODAY Published 1:27 p.m. ET Jan. 18, 2017 | Updated 7:44 p.m. ET Jan. 18, 2017

SAN FRANCISCO — Oracle is being sued by the Labor Department for allegedly paying white men more than their counterparts and for favoring Asian workers when recruiting and hiring for technical roles.

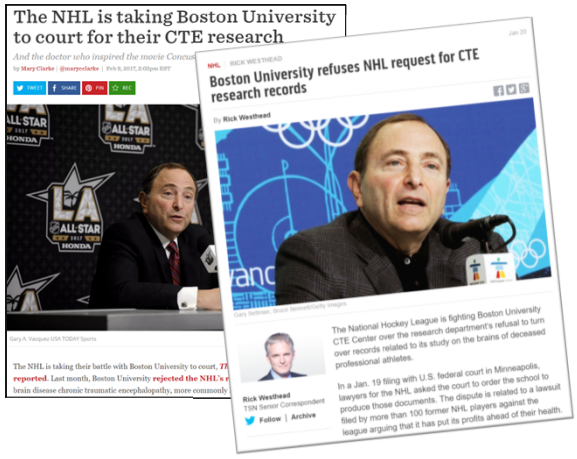
The administrative lawsuit is the latest from the

"In a statement, Google said it balked at turning over the private information of employees."

Can DOL prove (non)compliance without access to sensitive employee records?

The National Hockey League Question

Want to know if donated brains for CTE study belong to a cohort of players


The NHL is taking Boston University to court for their CTE research

Boston University refuses NHL request for CTE research records

By Rick Westhead

The National Hockey League is fighting Boston University CTE Center over the research department's refusal to turn over records related to its study on the brains of deceased professional athletes.

In a Jan. 19 filing with U.S. federal court in Minneapolis, lawyers for the NHL asked the court to order the school to produce those documents. The dispute is related to a lawsuit filed by more than 100 former NHL players against the league arguing that it has put its profits ahead of their health.

"BU objects to the production of documents concerning the study of the brains of hockey players whose families declined to authorize the release of such information or [those] whose participation was conditioned upon assurances of confidentiality."

BU letter to NHL, 10/26/2015

Can the court get an answer to NHL query without forcing BU to break its promise?

The Admission Racial Bias Question

Want to know if consideration of race in admissions results in reverse racism


THE WALL STREET JOURNAL.

SUBSCRIBE NOW | SIGN IN

GOP's Late Changes to Tax Bill Buoy Key Industries

Higher Education Act Proposal Primes Fight Over Future of Colleges

U.S. | EDUCATION

Harvard Faces DOJ Probe Over Affirmative-Action Policies

Justice Department accuses university of failing to cooperate in investigation of whether its admission policies discriminate against Asian-Americans

By Melissa Korn and Nicole Hong

Updated Nov. 21, 2017 3:12 p.m. ET


Slate NOV. 22 2017 3:11 PM

Justice Department Is Investigating Potential Racial Bias in Harvard's Admission Practices

By Lila Thulin

Can the DoJ get an answer without forcing Harvard to release admission data?

The answer to all these questions is **YES**

We can derive knowledge (K) from data $(x_1, x_2, x_3, \ldots)$ without requiring owners of the data to share it or to trust anything other than mathematics under some assumptions about threats

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

A True Story

July 31, 2014


Meeting with Mayor Menino @ BU, July 31, 2014


Katharine Lusk


Boston University Rafik B. Hariri Institute for Computing and Computational Science & Engineering

FOR IMMEDIATE RELEASE

CONTACT: Kira Jastive, 617-358-1240 or kjastive@bu.edu

(Boston) – Boston University's Rafik B. Hariri Institute for Computing and Computational Science & Engineering today announced it has received funding from the National Science Foundation (NSF) to develop a "smart-city" cloud platform designed to streamline and strengthen multiple municipal functions. Called *SCOPE: A Smart-city Cloud-based Open Platform & Eco-system*, the project is designed to improve transportation, energy, public safety, asset management, and social services in the City of Boston and across Massachusetts.

April 9, 2013

WOMEN'S WORKFORCE COUNCIL

The Women's Workforce Council was established by Mayor Thomas M. Menino on April 9th, 2013— known nationwide as Equal Pay Day. The day marks how far into 2013 women need to work to earn what men earned in 2012. The first of its kind in the country, the Council's mission is to help transform Boston into the best city in the country for working women.

Members of the Council represent the financial, engineering, medical, law, technology and retail sectors, and include small business owners, entrepreneurs, senior executives, as well as academic, labor and nonprofit leaders.

December 11, 2013

100% TALENT

The Boston Women's Compact

To make Greater Boston the premier place for working 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 workforce where all talent is cultivated 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.

December 11, 2013

100% TAL

The Boston Wom

To make Greater Boston the premier place for wor
closing the wage gap and removing the visible an
advancement. By doing so, we will build a more e
talent is cultivated and valued.

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.

GOAL 3

Evaluating Success

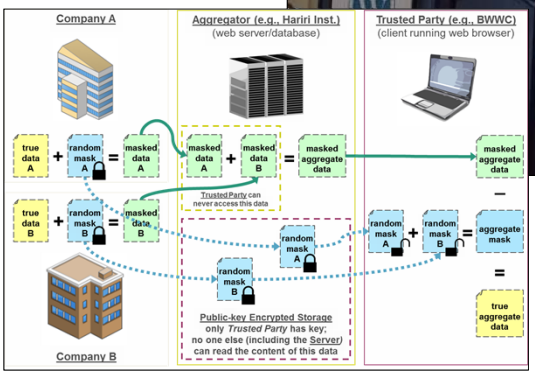
Employers agree to participate in a biennial review to discuss successes and challenges, as well as contribute data to a report compiled by a third-party on the Compact's success to date. Employer-level data would not be identified in the report. The specific data to be reported will build on data already required by federal and state authorities and should not create an additional reporting burden.

September 4, 2014 ++

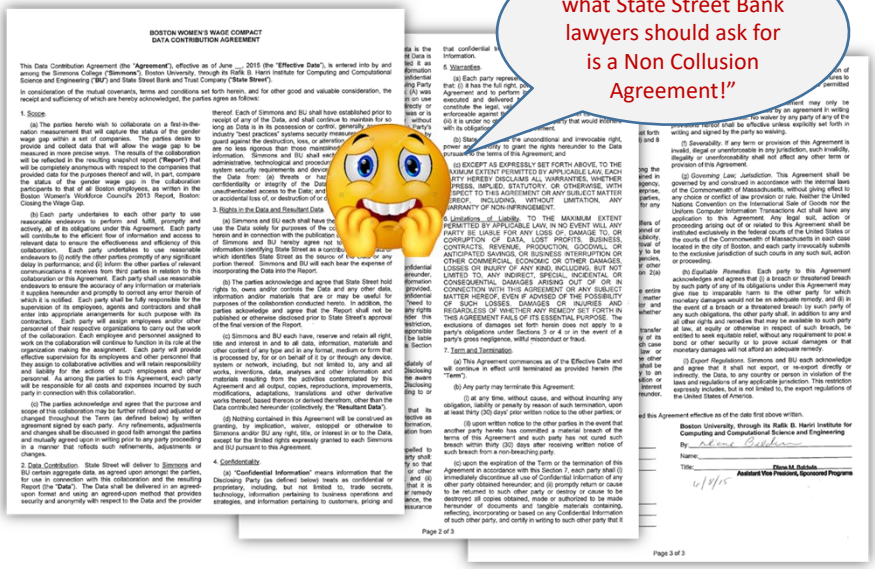
A subset of over 120 hours of meetings from Azer's Exchange Calendar with BWWC principals, Company CIOs, HR Officers, …

Subject	Start	Duration
Cathy Minehan	Fri 9/5/2014 10:30 AM	2 hours
Simmons College	Mon 10/27/2014 3:30 PM	1.5 hours
Data Collection for Pay Equity	Tue 12/2/2014 11:30 AM	30 minutes
Simmons College people	Fri 1/23/2015 1:00 PM	30 minutes
Invitation: 100% Talent Discussion with Data Partners @ Tue Mar 17, 2015 2pm - 3pm (johnstk3@s...	Tue 3/17/2015 2:00 PM	1 hour
Updated Invitation: MassMutual call with Hariri Institute re: Data Collection... @ Thu May 14, 2015 ...	Thu 5/14/2015 3:00 PM	1 hour
Invitation: Mock collection #1 @ Tue May 19, 2015 11am - 12pm (johnstk3@simmons.edu)	Tue 5/19/2015 11:00 AM	1 hour
Invitation: Mock Collection #2 @ Tue May 26, 2015 11am - 12pm (johnstk3@simmons.edu)	Tue 5/26/2015 11:00 AM	1 hour
Invitation: Mock Collection #3 @ Thu May 28, 2015 11am - 12pm (johnstk3@simmons.edu)	Thu 5/28/2015 11:00 AM	1 hour
Invitation: Call with BWWC @ Wed Jun 3, 2015 11:30am - 12pm (johnstk3@simmons.edu)	Wed 6/3/2015 11:30 AM	30 minutes
Updated Invitation: 100% Talent Data Collection: Hariri and Raytheon @ Fri Jun 5, 2015 9am - 10a...	Fri 6/5/2015 9:00 AM	1 hour
Invitation: 100% TALENT DATA COLLECTION @ Mon Jun 8, 2015 9am - 10:30am (johnstk3@simm...	Mon 6/8/2015 9:00 AM	1.5 hours
Invitation: Meeting with Boston Women's Workforce Council @ Tue Aug 11, 2015 10am - 11am (jo...	Tue 8/11/2015 10:00 AM	1 hour

April 14, 2015

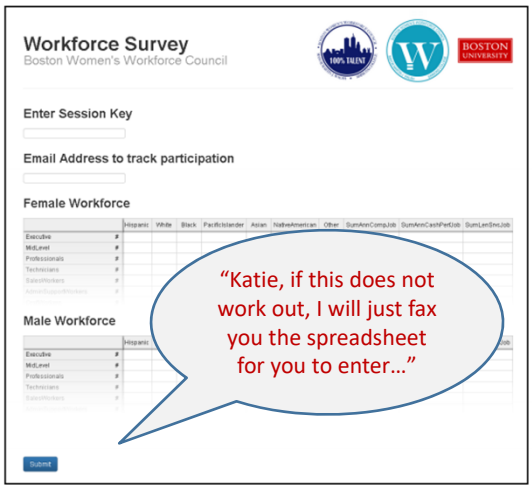


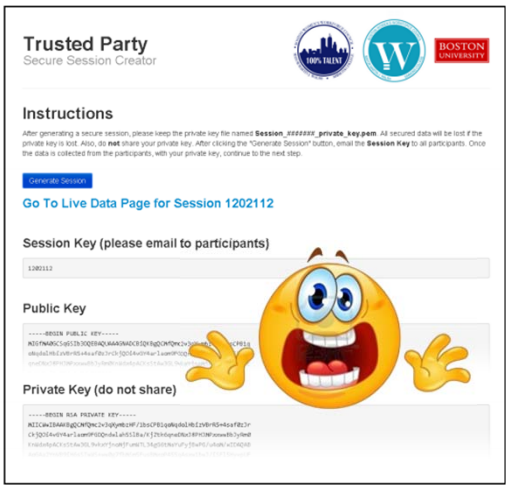
June 6, 2015



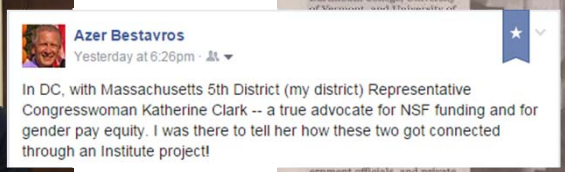
"Dennis and Diane, what State Street Bank lawyers should ask for is a Non Collusion Agreement!"

June 8, 2015 (D-day)





September 29, 2015



Scientists learn to reach audiences

SCIENTISTS
Continued from Page A1

Dartmouth College, University

ernment officials, and private philanthropists about the societal benefits of investing in research.
"We are in fiscally restrained

DREW ANGERER FOR THE BOSTON GLOBE
Computer science professor Azer Bestavros explained his work to Representative Katherine Clark in Washington.

On a recent morning, Bestavros and a Washington lobbyist for BU sat in the Capitol Hill office of Representative Katherine Clark, a Massachusetts Democrat on the House committee that oversees funding for the National Science Foundation.

Bestavros wanted to thank her for pushing for recent funding, impress upon her the importance of continued support, and arm Clark with ammunition for future budget and policy fights.

"I don't want to bore you with the details of what we do," he said, before launching into a story of how he uses cloud com-

The Boston Globe

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 Boston Globe

More Boston businesses join drive to end gender wage gap

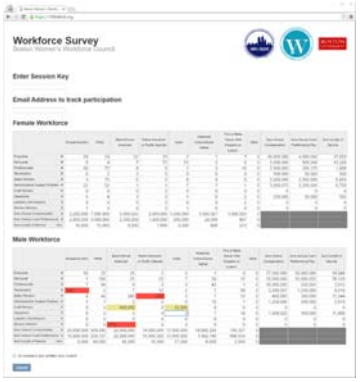
100% TALENT
The Boston Women's Compact

The Sequel

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

BU Today — In the World

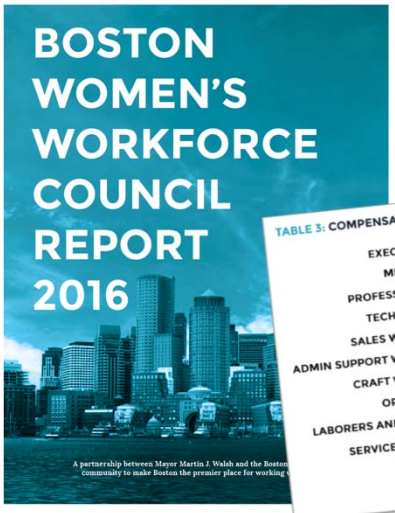
Calculating Gender Pay Equity
BU computer scientists remove obstacle in Boston's push for wage parity
07.08.2016 By Andrew Thurston



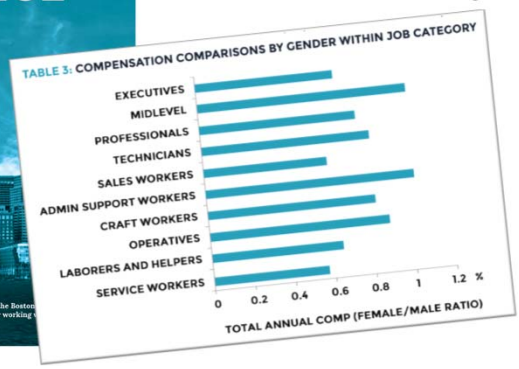
npr wbur 90.9 ON POINT WITH TOM ASHBROOK

BWWC co-chair Evelyn Murphy on secure multi-party computation: *"It's used in computer science applications, but it has never been used for public good. Here, we're beginning to show how to use this sophisticated computer science research for public programs."*

January 5, 2017



"We collected data regarding 112,600 employees, which represents 11% of the Greater Boston workforce and almost $11 billion in annual earnings."



Subscribe Starting at 99 cents | Members Sign In

Even in liberal Boston, there's a gender wage gap

ROBYN BECK/AFP/GETTY IMAGES

By Katie Johnston

GLOBE STAFF JANUARY 05, 2017

Working women in Greater Boston make 77 cents on the dollar compared to men — a gender wage gap that echoes the national average — according to a report released Thursday by the Boston Women's Workforce Council.

January 27, 2017

Boston Women Workforce Council @ Boston University

BU Today

Tackling the Wage Gap with Code

Hariri software team aids Boston Women's Workforce Council

Hariri Institute director Azer Bestavros at a City Hall press conference on the Boston Women's Workforce Council wage gap report earlier this month, flanked by Katie Conboy, Simmons College provost (from left), former Lt. Governor Evelyn Murphy, Boston Mayor Martin J. Walsh, and Eddie Ahmed, MassMutual Financial Group chief human resources officer. Photo by Cydney Scott

Mayor Walsh with members of the Boston Women's Workforce Council and 100% Talent Compact signers. Photo courtesy of the City of Boston.

CONTACT US:

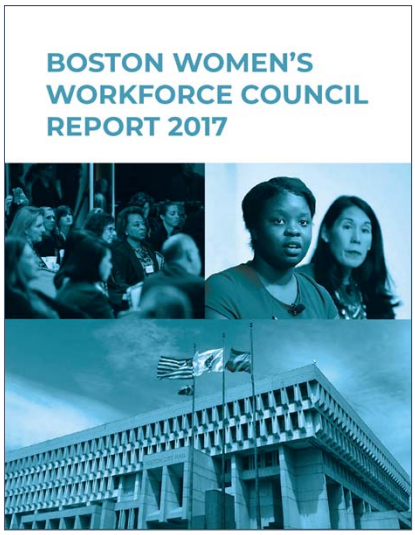
Boston Women's Workforce Council
Boston University Hariri Institute for Computing
111 Cummington Mall
Boston, MA 02215

Executive Director:

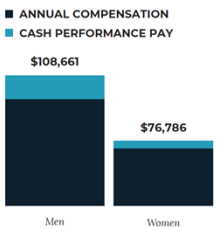
MaryRose Mazzola
maryrose.mazzola@bostonwomensworkforcecouncil.org

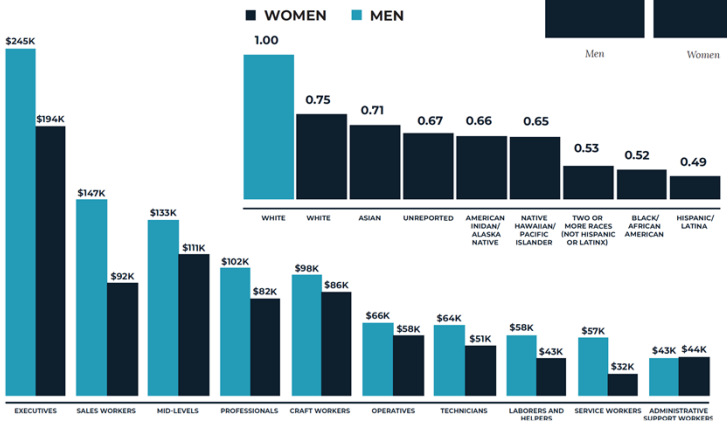
(617) 358-8517

January 31, 2018



"This year's data submission included 166,705 employees from 114 Compact-signing companies.





November 16, 2017

115TH CONGRESS
1ST SESSION

H. R. 4174

IN THE SENATE OF THE UNITED STATES
NOVEMBER 16, 2017
Received; read twice and referred to the Committee on Homeland Security and Governmental Affairs

AN ACT

To amend titles 5 and 44, United States Code, to require Federal evaluation activities, improve Federal data management, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.
(a) SHORT TITLE.—This Act may be cited as the "Foundations for Evidence-Based Policymaking Act of 2017".

GOVTECH WORKS by General Dynamics IT

Evidence-Based Policy Act Could Change How Feds Use, Share Data

by Tobias Naegele | Nov 29, 2017

As government CIOs try to get their arms around how the Modernizing Government Technology (MGT) Act will affect their lives and programs, the next big IT measure to hit Congress is coming into focus: House Speaker Paul Ryan's (R-Wis.) "Foundations for Evidence-Based Policymaking Act of 2017."

A bipartisan measure now pending in both the House and Senate, the bill has profound implications for how federal agencies manage and organize data – the keys to being able to put data for informed policy decisions into the public domain in the future. Sponsored by Ryan in the House and by Sen. Patty Murray (D-Wash.) in the Senate, the measure would:

November 29, 2017

115TH CONGRESS
1ST SESSION

S. _____

IN THE SENATE OF THE UNITED STATES

Mr. WYDEN (for himself, Mr. RUBIO, and Mr. WARNER) introduced the following bill; which was read twice and referred to the Committee on

A BILL

To establish a new higher education data system to allow for more accurate, complete, and secure data on student retention, graduation, and earnings outcomes, at all levels of postsecondary enrollment, and for other purposes.

1 *Be it enacted by the Senate and House of Representa-*
2 *tives of the United States of America in Congress assembled,*

Home > News > Press Releases

Wyden, Rubio, Warner Introduce "Student Right to Know Before You Go Act" to Empower Students as Consumers and Showcase New Privacy-Protecting Technology

Updated Legislation Allows Students and Families to Make Informed Decisions about How to Spend Their Higher Education Dollars; Protects Student Privacy By Featuring Encrypted, Secure Multi-party Computation

"We are excited to see legislation promoting the use of multi-party computation (MPC) in formulating sound public policy. Boston University's successful collaboration with the City of Boston and the Boston Women's Workforce Council brought this technology into practice to maintain data privacy while gaining insight into an important societal issue -- potential wage inequality in private industry. Such applications demonstrate that MPC can bring enormous value to policymakers at all levels of government."

-- Azer Bestavros (on behalf of the team from BU)

August 21, 2018

"the bills have been stalled in the Senate, caught in the tension between two constituencies that mean well. One wants to improve consumer awareness and one wants to protect the privacy of individuals in an age when there isn't much privacy left. It's an increasingly familiar conflict between privacy and the public good, and because of a little-used cryptographic technology that has been used to solve similar clashes, it need not exist."

-- Azer Bestavros
Washington Post Op-Ed, August 21, 2018

The Washington Post
Democracy Dies in Darkness

Grade Point • Perspective

It's time to tell students what they need to know

By Azer Bestavros
August 21

What's the real value of a college education? For some people, that is calculated by the earning power that comes with a particular college degree. For others, it's more about the satisfaction of a particular career, or about the social skills or intellectual rewards gained from the college experience. In every case, it is a life-altering decision with little data to inform it. Less data, in fact, than many other types of major long-term expenses and commitments, such as purchasing a home or buying a car.

Prospective students and their parents deserve a better look at what they are signing on for, and they could get exactly that with either of two bipartisan bills in the Senate. The Wyden-Rubio Student Right to Know Before You Go Act and the College Transparency Act would require colleges to collect and share data about such things as college costs, post-graduation salaries, the percentage of students that go on to graduate school and how long it takes to graduate. The initiatives would give prospective students a great deal of information that could help many decide what college to attend.

No one thinks that's a bad idea, but many people — privacy advocates, congressional Republicans and private colleges — point out that the collection and dissemination of such student-level data could violate a ban written into the 2008 reauthorization of the Higher Education Act. Consequently, the bills have been stalled in the Senate, caught in the tension between two constituencies that mean well. One wants to improve consumer awareness and one wants to protect the privacy of individuals in an age when there isn't much privacy left. It's an increasingly familiar conflict between privacy and the public good, and because of a little-used cryptographic technology that has been used to solve similar clashes, it need not exist.

Multi-Party Computation (MPC)

What is it?

- Given multiple parties $p_1, p_2, \ldots, p_n$ each with private data $x_1, x_2, \ldots, x_n$
- Parties engage in computing a function $f(x_1, x_2, \ldots, x_n)$
- Nothing is revealed about the inputs beyond what the output of f reveals
- What f leaks is an orthogonal question, e.g., the realm of "differential privacy"

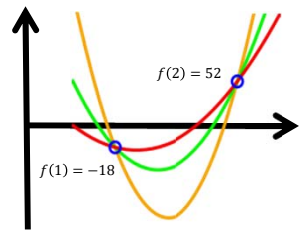
State of the Art

- Theory known since 1979, with Shamir's "How to share a secret"
- Frameworks and libraries increasingly available over the last few years …
- Experience with real use cases at scale is limited ← We are changing that
- Deployments are not easily portable ← We are changing that

Shamir Secret Sharing (1979): The Basic Math

➔ Need $k+1$ points to define polynomial of degree k

- To share a "secret" among k parties, make it the free coefficient of a polynomial $f(x)$ of degree k
- Select coefficients $a_1, a_2, \ldots, a_k$ of $f(x)$ at random
- Give party P_i a "share" of the secret – namely, $f(i)$
- To reconstruct the "secret" all parties need to combine their shares to find the secret – namely $f(0)$



$$\boldsymbol{f(x) = 32 - 110x + 60x^2}$$

a_1 a_2

Notes

- Need to use finite field arithmetic to provably avoid any leakage
- Approach allows secret sharing among any number of parties; any subset k can uncover the secret
- Other approaches have been proposed, most notably the use of garbled circuits

Multiparty Computing on Secret Shares

Any arbitrary function is a circuit of additions & multiplications

➔ Addition is easy!

- Sum of secrets is represented by $f(x) = f_1(x) + f_2(x)$
- To compute $f(x)$, each party adds its shares of $f_1(x)$ and $f_2(x)$
- Using one round of k messages, sum of secrets can be revealed

➔ Multiplication is not that easy…

- Multiplication of secrets is represented by $f(x) = f_1(x) * f_2(x)$
- Requires $O(k)$ rounds of communications – could be very expensive!

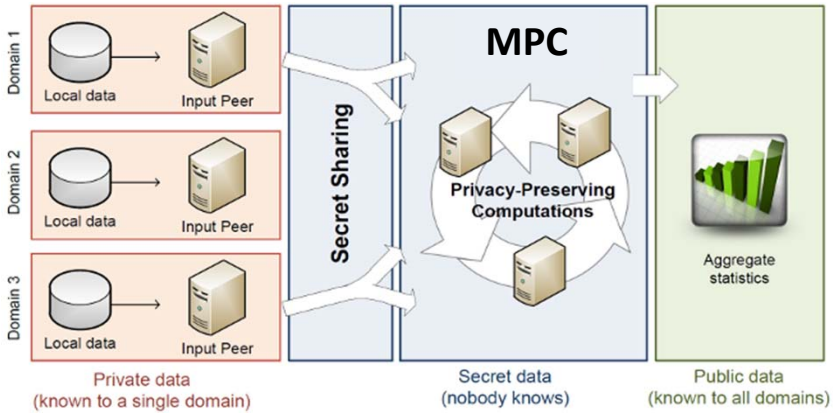
Another Flavor: Yao Garbled Circuits (1986)

- Motivated by Yao's Millionaires Problem (who is wealthier)
- Enables two mistrusting parties to jointly evaluate a function over private inputs using "oblivious transfer" (OT) primitive
  - P_1 replaces inputs of a truth table (gate in circuit) with random labels
  - P_1 encrypts truth table outputs using corresponding input labels
  - P_1 permutes the table and sends the encrypted "garbeled" table to P_2
  - P_1 sends the labels corresponding to its private input to P_2
  - P_1 also sends the labels corresponding to P_2's inputs to P_2 using OT
  - P_2 uses labels corresponding to private inputs to compute output label
  - P_2 communicates output label to P_1 who decrypts it and reveals result

Secret Sharing: How?

$$f(x) = S + r_1 x^1 + r_2 x^2 + r_3 x^3 + \ldots + r_i\, x^i + \ldots$$

$$s_1 = f(1)$$
$$s_2 = f(2)$$
$$s_3 = f(3)$$
$$\ldots = \ldots\ldots$$
$$s_i = f(i)$$
$$\ldots = \ldots\ldots$$

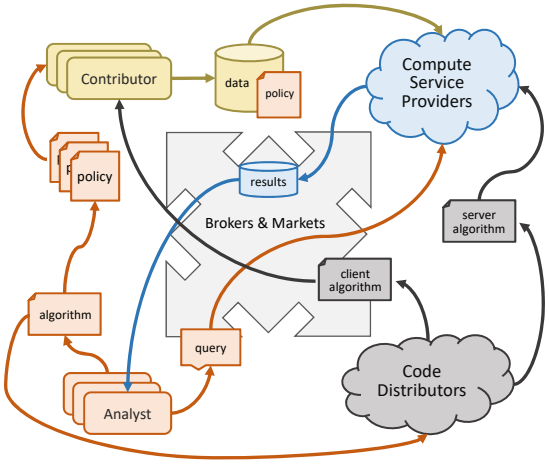


Data Marketplace: Stakeholders & Interfaces

Contributor
- Makes data available
- Sets/chooses policies

Broker or Market
- Identifies data contributors and analysts that have mutually compatible analysis algorithms/data policies
- Helps data contributors craft or choose from possible policies, and helps analysts choose algorithms

Data Analyst
- Defines analysis algorithms
- Chooses performance criteria
- Examines results



Compute Service Provider
- Ingests and stores secret-shared data
- Performs MPC computations
- Optimizes where policies allow

Code Distributor
- Distributes code to contributors and/or service providers
- Provides assurance or auditing that code is from the right analyst

Modeling threats and adversaries

Crypto MPC researchers consider four types of adversaries

- Semi-honest adversary:
  - Follows rules but may attempt to glean information along the way
- Covert adversary:
  - Cheats only if unlikely to be caught
- Rational adversary:
  - Cheats as long as expected payout is larger than expected penalty if caught
- Malicious adversary:
  - Performs any action needed to breach system integrity

The Parties in our MPC Setting

Contributors (100% Talent Companies)

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

Broker + Analyzer (BWWC)

- Ultimate recipient of the output of the analytic
- May also participate in computing the analytic

Service Provider + Code Distributor (BU)

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

Threat Modeling & 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

… but place no trust in service provider

- Service provider cannot be entrusted with data or with the results
- Assume that service provider is incentivized to perform the computation on behalf of the contributors and analyzers

Multi Party Computation: State of the Art

Very active R&D to make MPC accessible to programmers:

Frameworks

- ABY - 2PC with secret sharing and GC; semi-honest adversaries
- batchDualEx - 2PC with GC; malicious adversaries
- Duplo - 2PC GC; malicious adversaries
- Obliv-C - 2PC with gGC; semi-honest adversaries
- Sharemind - 2PC or 3PC with secret sharing; semi-honest adversaries
- SPDZ - General MPC with secret sharing; malicious adversaries
- TinyLEGO - 2PC with GC; malicious adversaries
- Viff - General MPC with secret sharing; semi-honest adversaries

Tools

- CBMC-GC - Creates Boolean circuits (GC) from ANSI-C code
- UC Compiler - Valiant's Universal Circuit Compiler

Primitives

- APRICOT - OT Extension secure against malicious adversaries
- libOTe - Library with various OT Extensions.
- OT Extension - OT Extension secure against malicious adversaries
- SCAPI - Various secure computation API's
- SplitCommit - Additively homomorphic commitment scheme
- TSS - Pure-Rust implementation of threshold secret sharing schemes

Protocols

- BaRK-OPRF - Private Set Intersection
- Linreg - Privacy preserving linear regression
- ORAM (Obliv-C) - Oblivous RAM
- PSI - Private Set Intersection

Commentary on State of Art

Adversarial models are too simplistic

- Need to match crypto threat models with economic, reputation, and legal incentives
- Design of privacy-preserving platforms should take advantage of more realistic models
- Plausible deniability (e.g., participation in MPC) goes beyond keeping data private
- Need to account for the weakest link – the human in the loop!

All parties are not created equal

- Parties may have significantly different backend systems and technical sophistication
- Parties interested in output of MPC may not be the owner of the private data
- Privacy concerns are not uniform across all parties

➔ **Need to design solutions that match stakeholders & roles**

Research Projects @ Boston University

Develop new MPC primitives, toolkits, and optimizations

- Efficient shortest-path algorithms operating over private subgraphs
- Efficient analytics/personalization over private geo-temporal data
- PL and compiler frameworks to expose privacy-utility tradeoffs

Develop MPC "as a service" solutions in various settings

- Web/browser-based MPC as a service platform
- Spark-based MPC platform for Map-Reduce analytics
- Incorporate MPC in big-data cloud workflow management

Open-source MPC Libraries

JIFF: JavaScript Implementation of Federated Functionalities
Library for building web-based applications using secure multi-party computation
https://github.com/multiparty/jiff

Web-MPC
JavaScript application for user-friendly privacy-preserving web-based data aggregation
https://github.com/multiparty/web-mpc

Conclave Workflow Manager
Compiler that optimizes relational queries to be executed under MPC by factoring it into (1) scalable, local, cleartext processing workflows using backends such as Apache Spark, and (2) isolated MPC workflows that utilize existing MPC backend frameworks
https://github.com/multiparty/conclave

MPC as a Service – killer apps…

Systemic Threat Analytics in Federated Settings

- Banking and Finance
- Data Network Operations

Collective Intelligence in Competitive Settings

- Information Brokerage for Business/Marketing Intelligence
- E-Commerce Analytics over Segmented Proprietary Data Assets
- Personalization and Sharing Economy Applications

Public Good Settings

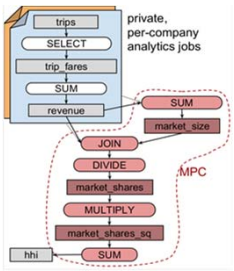
- Privacy-preserving Sensus and Surveys
- Healthcare, Education, and Academic Research
- Compliance Testing/Reporting for Trade Associations
- Private/Fair Reporting of Sexual Harrasement/Abuse in Workplace

ACSC Advanced Cyber Security Center
LexisNexis
Boston Medical Center
HRI Honda Research Institute EU
MBA Massachusetts Bankers Association
OCA Massachusetts Office of the Child Advocate

MPC for big-data cloud workflow management

Our Solution

- **SQL-like DSL Programming**
  - ➔ *No MPC experience necessary*
  - ➔ *Separate InfoTech from InfoSec*
- **Compiler does MPC transforms**
  - ➔ *No need for privacy experts*
  - ➔ *No need for systems experts*
- **Dispatcher for local deployment**
  - ➔ *No need for new backend*
  - ➔ *No cross-platform integration*



Herfindahl-Hirschman Index on 156GB NYC trip data

Setup	Runtime
Insecure, trusted Hadoop (8 nodes)	16 min 10 s (970s)
Musketeer with MPC (5 parties, 1+1+1+1+4 nodes)	17 min 31 s (1,051s)
Secure MPC framework only (VIFF only, 5 parties, 5 nodes)	>2 hours (7,200s)

Takeaway: We can have it both ways

We can derive knowledge (K) from data $(x_1, x_2, x_3, \dots)$ without requiring owners of the data to share it or to trust anything other than mathematics under some assumptions about threats

$$\mathrm{K} = f(\text{CONFIDENTIAL}, \text{CONFIDENTIAL}, \text{CONFIDENTIAL}, \dots)$$

When it comes to data and computation over data, we need to rethink our notions of ownership, custody, jurisdiction, sharing, disclosure, liability, and introduce new ones such as collusion.

Takeaway: Societal Implications

- Privacy/confidentiality concerns should not be used as excuses to deny society the right to answer important questions
- Privacy/confidentiality should not be sacrificed in the name of doing the right thing, or advancing science, or applying the law
- Private data should not be a tradable commodity; computation over private data should be what we offer “as a service”
- Substantial social/financial value can be gained in contexts imposing legal or policy restrictions on sharing raw data

Acknowledgments: It takes a village!

www.multiparty.org

Andrei Lapets | Kyle Holzinger | Eric Dunton | Frederick Jansen | Nikolaj Volgushev | Mayank Varia | Malte Schwarzkopf | Kinan Bab | Rawane Issa

BU The Rafik B. Hariri Institute for Computing and Computational Science & Engineering

Azer Bestavros: Sharing Knowledge without Sharing Data -- On the false choice between privacy and utility of information 43

Hariri Institute for Computing

"Leveraging the Computational Perspective in a Data-Driven World for a Better Society"

Website: www.bu.edu/HIC
Twitter: @BU_Computing
Facebook: BUcomputing

Azer Bestavros: Sharing Knowledge without Sharing Data -- On the false choice between privacy and utility of information 44