PEPR '20

Building and Deploying a Privacy Preserving Data Analysis Platform

Frederick Jansen || fjansen@bu.edu

Director – Software & Application Innovation Lab (sail.bu.edu)

CTO – Nth Party (nthparty.com)

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

Nth party

Software & Application Innovation Lab

BOSTON UNIVERSITY

What is MPC?

Private

Inputs

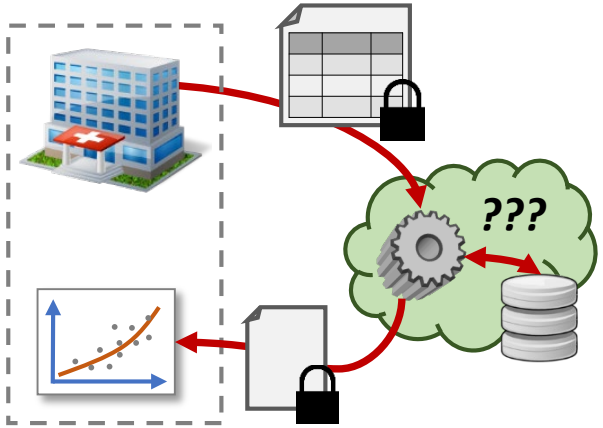
$$f(s_1, s_2, s_3, \ldots, s_n) = z$$

Public

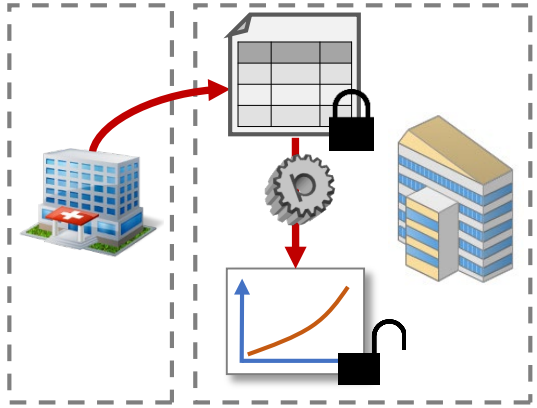
Output

Privacy-Enhancing Technologies

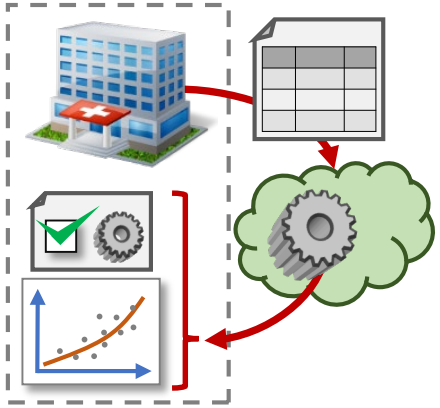
Homomorphic Encryption



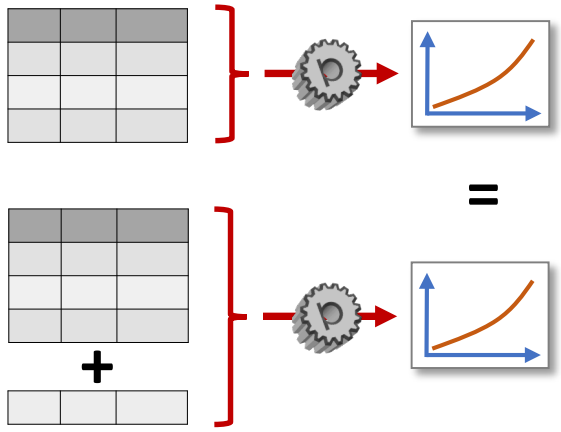
Functional Encryption



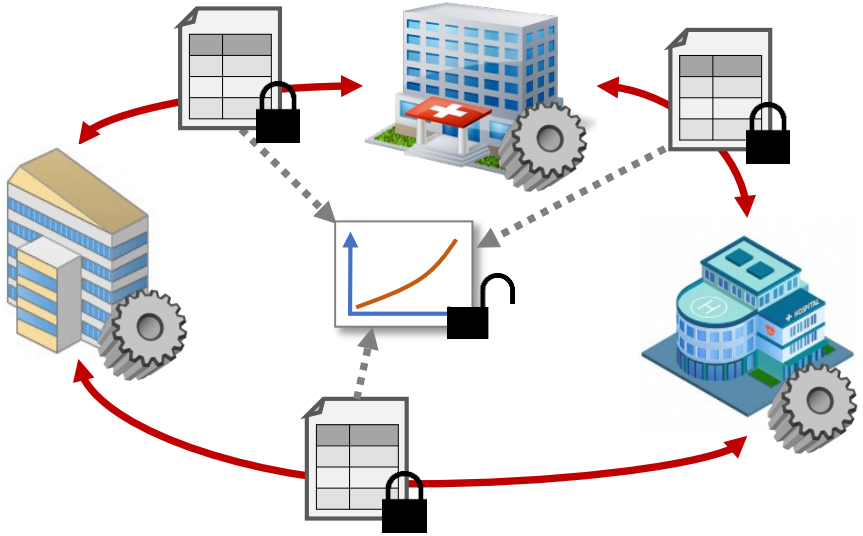
Zero-knowledge Proofs



Differential Privacy



Multi-Party Computation



MPC for Social Good

Hey,Charlie

Combatting Opioid Addiction

BWWC
BOSTON WOMEN'S WORKFORCE COUNCIL

Pay Equity

Bank Fraud

Greater Boston Chamber of Commerce

Economic Inclusion

The Dataverse® Project

Social Science Research Sharing

CALLISTO

Detecting Serial Perpetrators of Sexual Misconduct

Two years ago

Sugar Beet Auction

ALEXANDRA INSTITUTTET

AARHUS UNIVERSITY

DANISCO

Danske Sukkerroedyrkere

Salary Equity

BWWC
BOSTON WOMEN'S WORKFORCE COUNCIL

CITY of BOSTON

2008 | 2014 | 2015 | 2016 | 2017 | 2018

Tax Fraud Detection

CYBERNETICA

REPUBLIC OF ESTONIA
TAX AND CUSTOMS BOARD

Corporate Spending on Minority-Owned Businesses

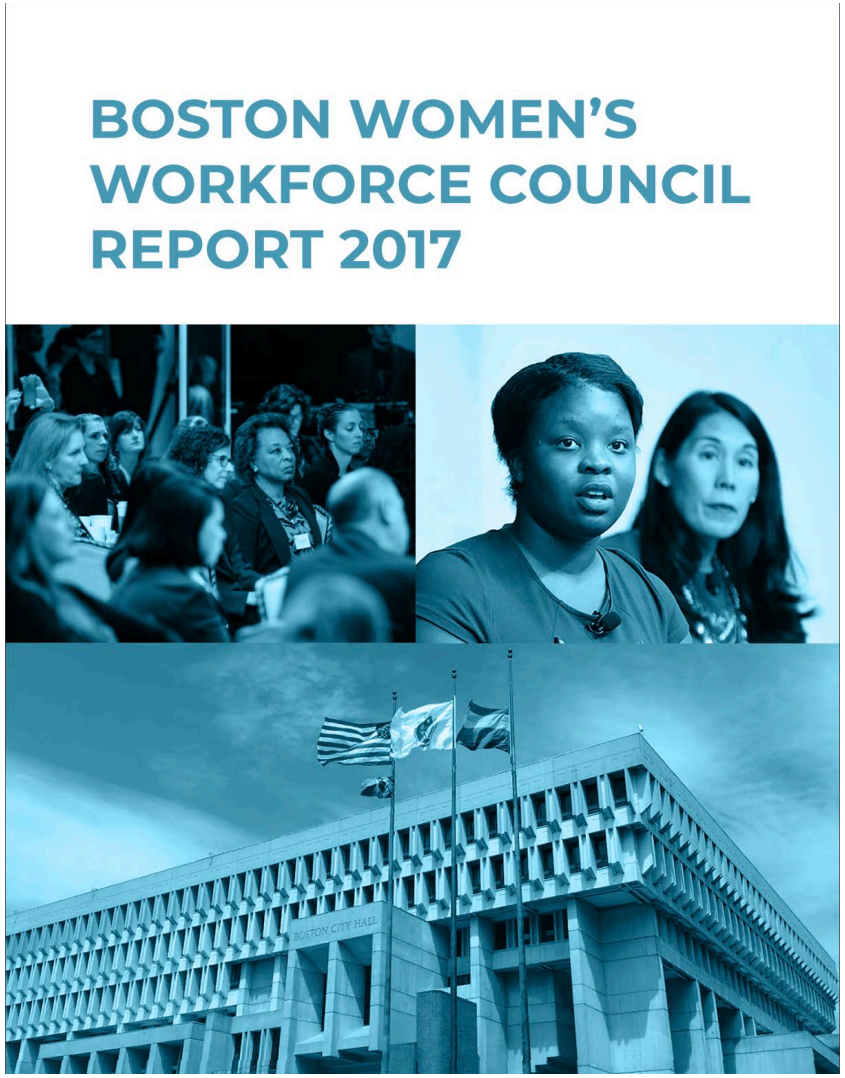
Greater Boston Chamber of Commerce

Now

ZenGo
UNBOUND (MATH OVER MATTER)
SEPIOR
CYBAVO
CYBERNETICA
enigma
US TECHNOLOGIES
keyTango
NuCypher
OFFCHAIN LABS
PARTISIA
curv
INSIGHTS NETWORK
PlatON
BYTOM
Alibaba Group
ATOMRIGS LAB
blocksecsys
galois
fragmentiX
SPHERITY
Fressets
DATA SECRETS
qedit
VANTEUM
AMIS
ARPA
cryptoworth
Acronis
SHARDX
XXLSEC
BOLT LABS
I4P
ITRI
cosmian
TSINGJIAO INFORMATION SCIENCE
SAMSUNG SDS
Digital Garage
nth party
MYKEY
Qredo
蚂蚁金服 ANT FINANCIAL
salesforce
GROUND X
Penta SECURITY
THRESHOLD
KEYLESS

MPC Alliance members

Setting the stage



"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."

BWWC
BOSTON WOMEN'S WORKFORCE COUNCIL

BOSTON WOMEN'S WORKFORCE COUNCIL
An Initiative of Mayor Martin J. Walsh to close the wage gap
In Partnership with Simmons College and Greater Boston businesses

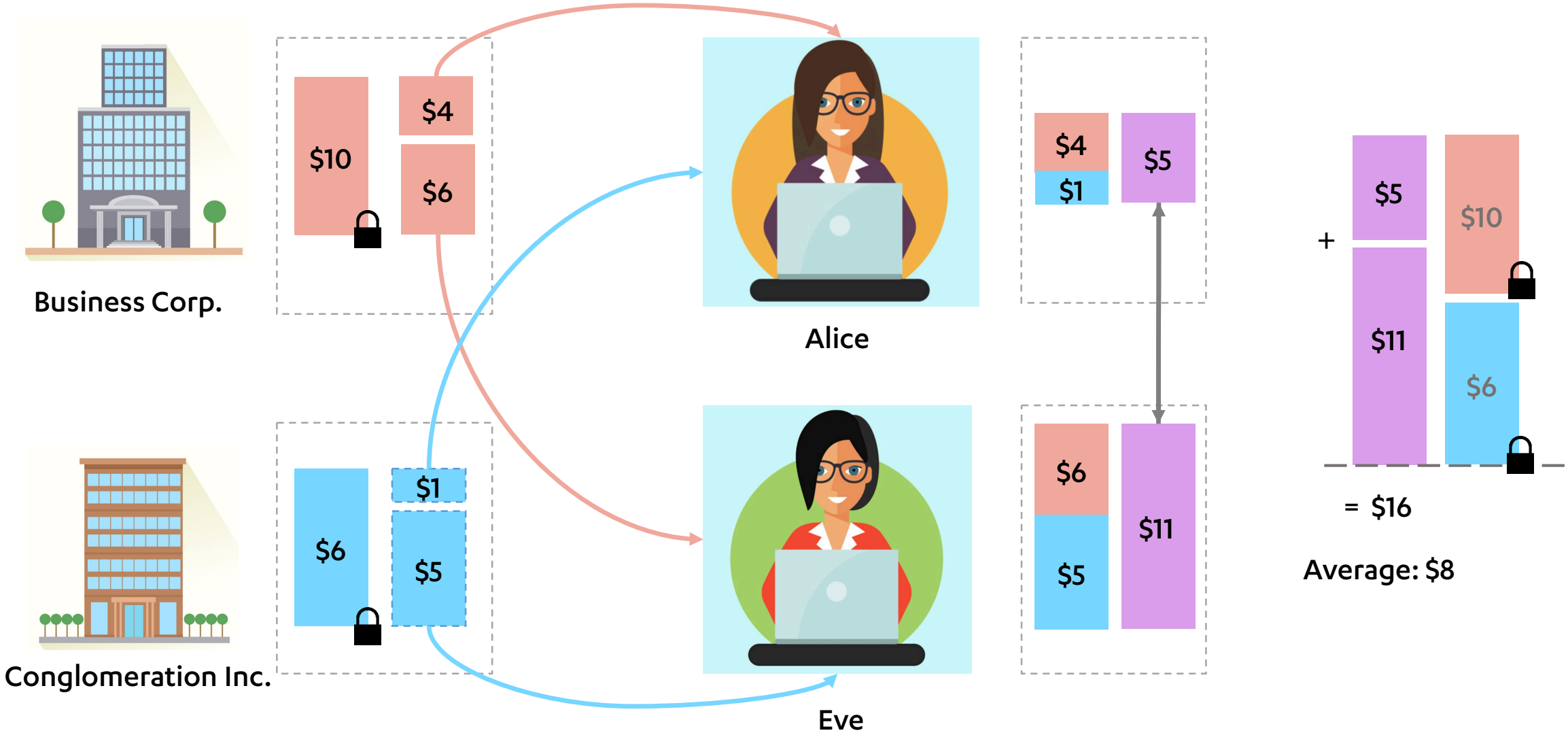
Goals:

- Understanding the root causes of the wage gap
- Closing the gap
- Evaluating success

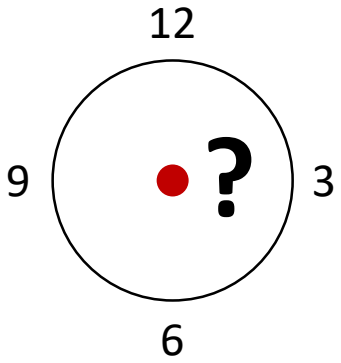
Communication is key

- Should I participate?
  - Access control, encryption at rest, privacy, bad press, trusted 3rd party, cost, expertise, …
- Can I participate?
  - HIPAA, FOIA, liability, NDA, …

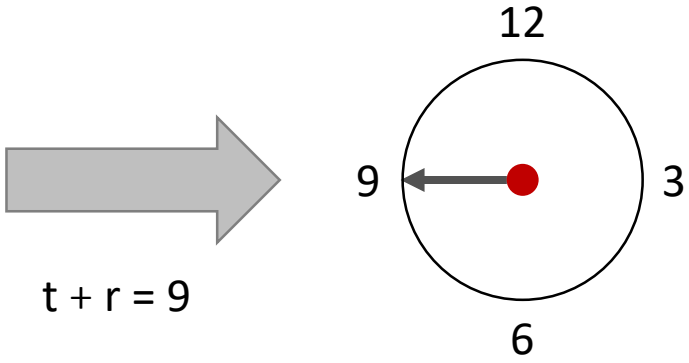
Additive Secret Sharing #1

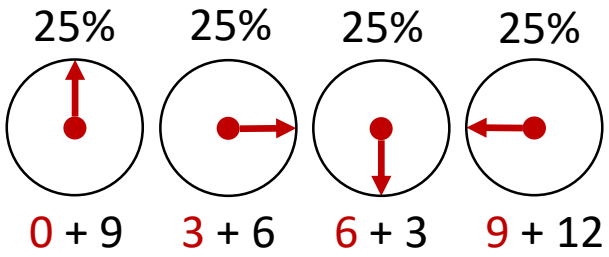


Clocks for Finite Field Arithmetic

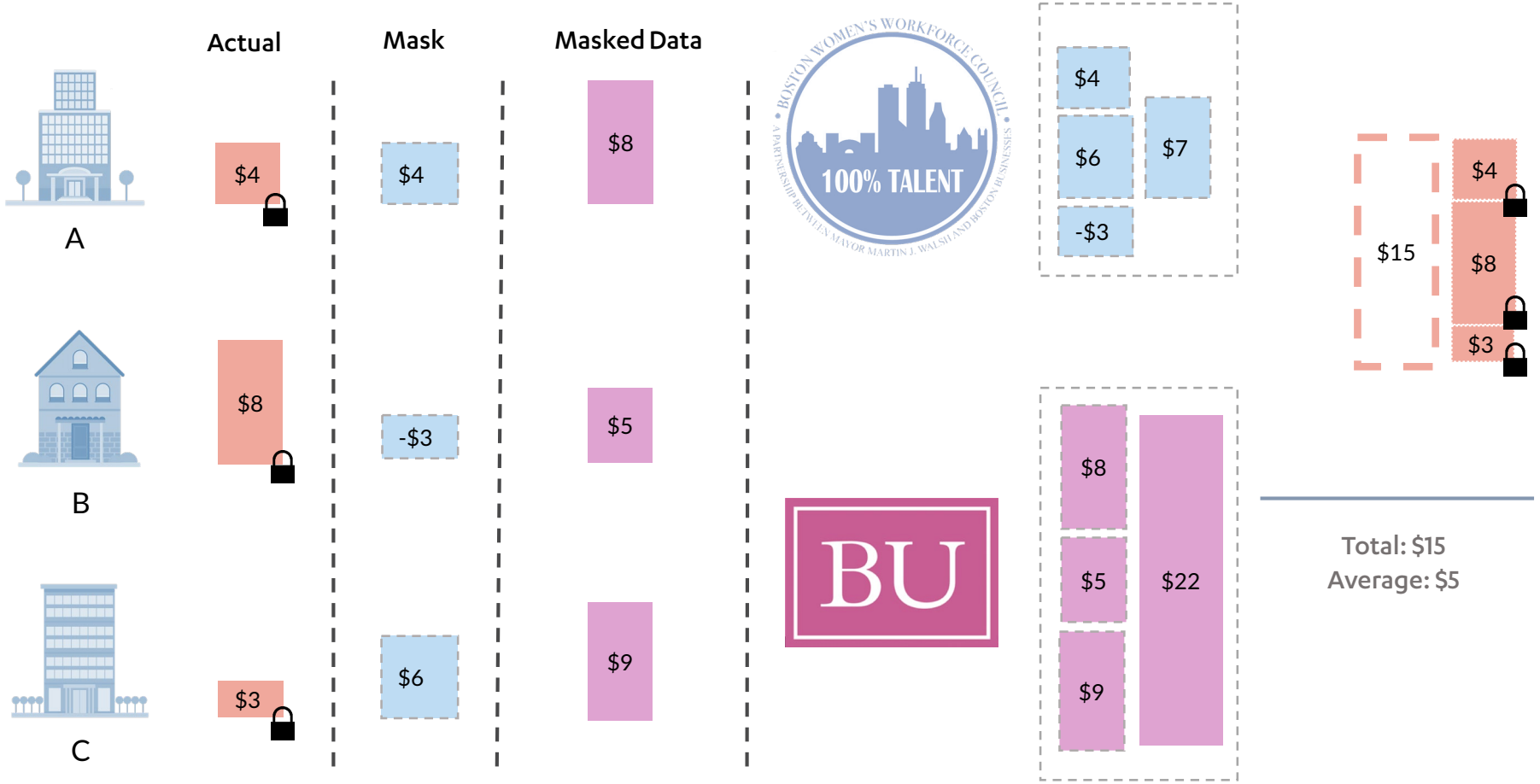


r	Prob.
+3	25%
+6	25%
+9	25%
+12	25%

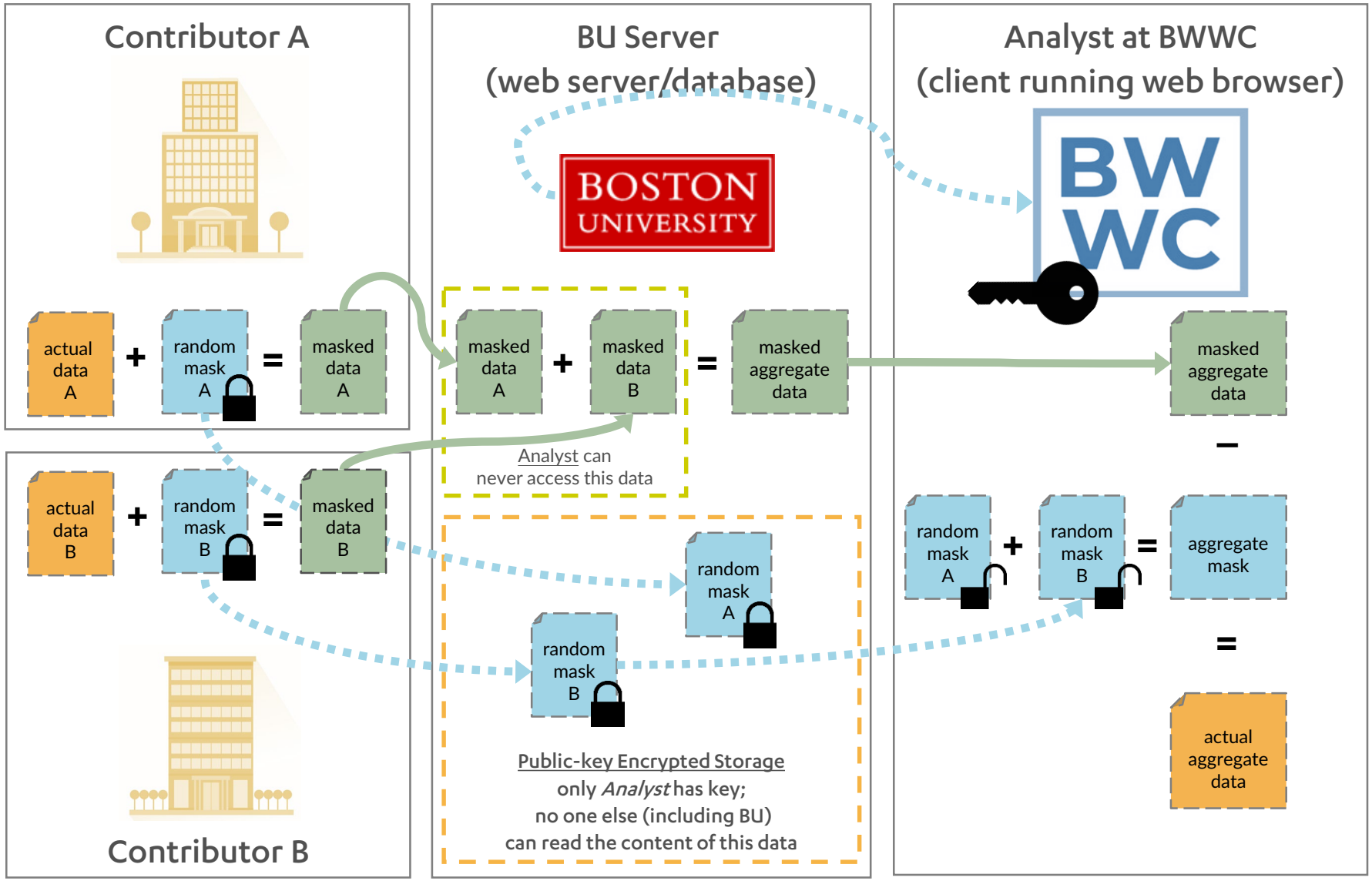




Additive Secret Sharing #2



Protocol diagram



Other considerations

- Extrapolate from here, trust us
- Build trust through participants and track record
- Other variants of presentations (Shamir's Secret Sharing) exist
- Average birthday example for in-person settings
- Training sessions and handouts for data submission
- Verbiage? Upload vs submit vs participate?

Why build anew?

- Focus on
  - Accessibility
  - Familiarity
  - Verifiability
  - Participation
  - Explainability

```
28          twopc.function_dependent();
29          io->flush();
30          cout << "dep:\t"<<party<<"\t"<<time_from(t1)<<endl;
31
32      // create and fill in input vectors (to all zeros with memset)
33          bool *in = new bool[max(cf.n1, cf.n2)];
34          cout << "input size: max " << cf.n1 << "\t" << cf.n2;
35          bool * out = new bool[cf.n3];
36          if (party == ALICE) {
37              memset(in, false, max(cf.n1, cf.n2));
38          } else {
39              memset(in, false, max(cf.n1, cf.n2));
40          }
41          memset(out, false, cf.n3);
42
43      // online protocol execution
44          t1 = clock_start();
45          twopc.online(in, out);
46          cout << "online:\t"<<party<<"\t"<<time_from(t1)<<endl;
47
48      // compare result to our hardcoded expected result
49          if(party == BOB and check_output.size() > 0){
50                  string res = "";
51                  for(int i = 0; i < cf.n3; ++i)
52                          res += (out[i]?"1":"0");
53                  cout << "result: " << res << endl;
54                  cout << (res == hex_to_binary(check_output)? "GOOD!":"BAD!")<<endl;
55          }
56          delete[] in;
```

Version #1

- Developed in 1 week by interns
- Worked, for the most part

Enter Session Key

153nk3qwhb39a1g56d89

Enter Participation Code

5sh9q9r2gk60xtk2bb447

Amount Spent with MBEs

	Value for FY in Thousands of Dollars ($)
Dollar Amount Spent with Local MBEs ($)	$11K
Dollar Amount Spent with State MBEs ($)	$52K
Dollar Amount Spent with National MBEs ($)	$23,000,000K

Addressable Spend

	Value for FY18 in Thousands of Dollars
Total Dollar Amount Spent Procuring All Goods and Services Locally ($)	
Total Dollar Amount Spent Procuring All Goods and Services at the State Level ($)	$39K
Total Dollar Amount Spent Procuring All Goods and Services in the United States ($)	$521K

Number of MBEs

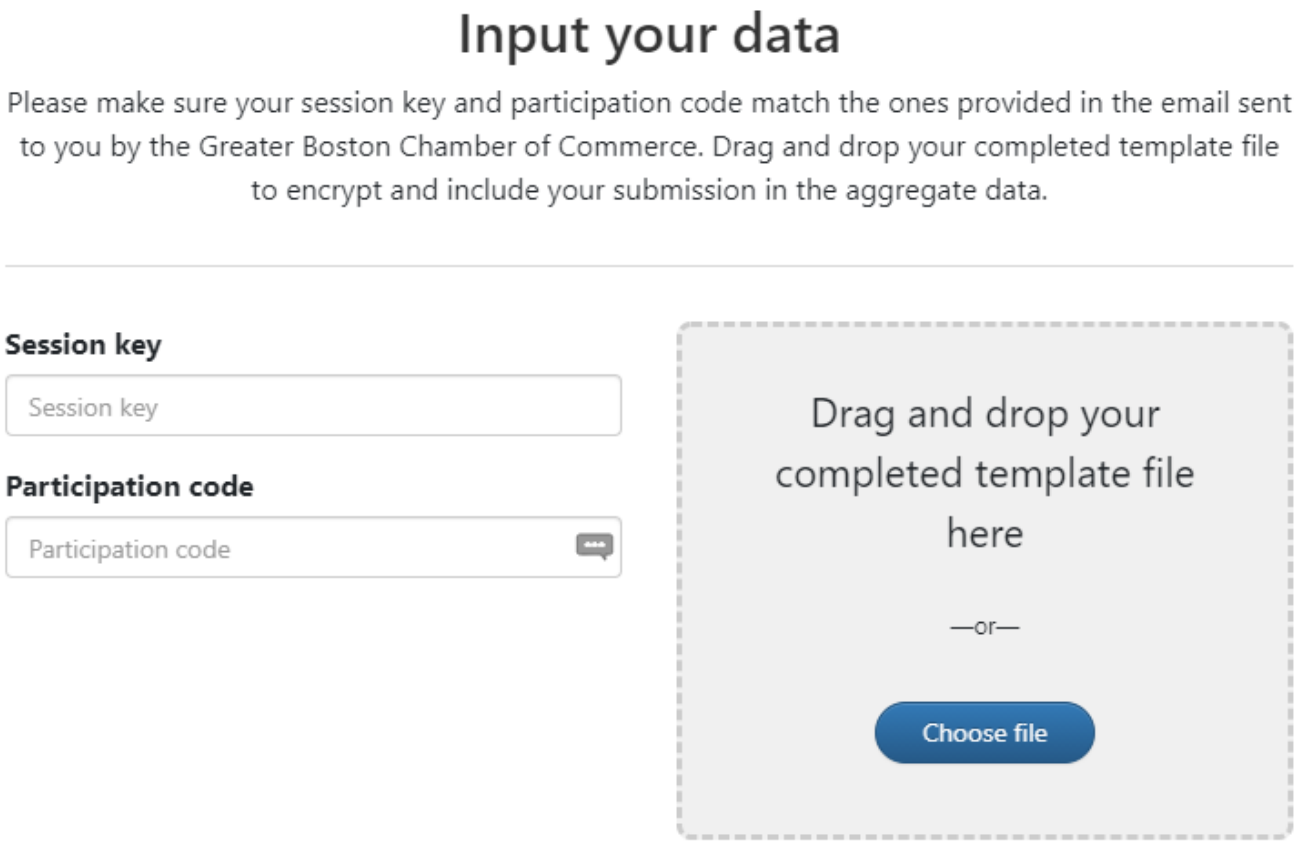
	Value for FY18
Number of Local MBEs With Whom You Have Done Business (#)	12
Number of State MBEs With Whom You Have Done Business (#)	56
Number of National MBEs With Whom You Have Done Business (#)	199

☐ All numbers are verified and correct

Submit

Version #3

- Result of extensive user testing
- Supported 100s of participants
- Configurable
- Drag-and-drop



Error checking

Asian		American Indian/Alaska Native		Two or More Races (Not Hispanic or Latinx)	
Female	Male	Female	Male	Female	Male
0	0	0	0	0	0
18	10000000	110	111	112	113
28	29				
38	39				
48	49	410	411	412	413

Warning: Data is too big

Are you sure this value is correct?

adfs	$47.00	$48.00	$49.00	$410.00	$411.0
$56.00					
$66.00					
$76.00	$77.00	$78.00	$79.00	$710.00	$711.0

Invalid Data Entry

Please do not input any text or leave any cells blank. If the value is zero, please input zero.

Verify and submit your data

Please ensure that all data entered is accurate, and confirm that all employees are accounted for by reviewing the total number of employees below.

Totals Check

	Total Number of Employees		
	Female	Male	All
Total	15905	16390	32295

☐ All data is verified and correct

Errors

- Invalid session number
- Invalid participation code
- Please answer all Additional Questions

Submission history

- You have not submitted yet

Submit

Analytics under MPC

Answer additional questions

We have included these questions to get instant feedback as to how this process went in order to improve the process in future years. Please know that the answers to these questions will be anonymous, and they will be considered separately from the encrypted and aggregated data above.

Which department are you in?

- Human Resources (e.g. HR Manager, HRIS Manager, Compensation Manager, Talent & Development)
- Operations (e.g. Director of Operations)
- Diversity (e.g. Chief Diversity Officer)
- Upper Management (e.g. COO, CEO, Executive Director)
- Other

What kind of HRIS or organizational system does your company/organization use?

- Large-scale traditional HRIS/HRMS software (e.g. ADP, Workday, PeopleSoft, etc.)
- Microsoft Office or similar (e.g. Excel, Microsoft Word, Google Docs)
- Other

How easy was it to understand what data was required given the template and instructions?

- Extremely easy
- Moderately easy
- Slightly easy
- Neither easy nor difficult
- Slightly difficult

- Slightly difficult
- Moderately difficult
- Extremely difficult

How long did it take to prepare the data for submission given your organizational system?

- Less than 1 business day
- 1-3 business days
- 4-7 business days
- 7-10 business days
- Greater than 10 business days

Verify and submit your data

Please ensure that all data entered is accurate, and confirm that all employees are accounted for by reviewing the total number of employees below.

Totals Check

	Total Number of Employees		
	Female	Male	All
Total	15905	16390	32295

Submission history

- You have not submitted yet

All data is verified and correct

Submit

Quantifying privacy and leakage

- Philosophical question more than technical one
- Some decisions are easy
  - No queries with single row result (maybe)
  - No repeated queries on minor data modifications
- Others not so much
  - How many participants needed?
  - What if participants are anonymous?
  - Which algorithms are appropriate?
- Differential privacy + MPC ideal (sometimes)

All the times we messed up

- Deployment 1: not enough data validation (invalid data)
- Deployment 2: not enough data validation (semantic errors)
- Best effort on browser support, participant asking to fax data
- Assuming wrong upper bound, leaking lower bound
- Recovering from data entry error under MPC, leaking order of magnitude

Acknowledgements

Azer Bestavros
Nikolaj Volgushev
Andrei Lapets
Ben Getchell
Ibrahim Shaikh
Kyle Holzinger
Wyatt Howe
Tanner Braun
Rachel Manzelli
Ran Canetti
Rawane Issa
Malte Schwarzkopf
Assaf Kfoury
Justin Chen
Eric Dunton
Ashni Shah
Mike Gajda
Ira Globus-Harris
Mayank Varia
Kinan Dak Albab
Lucy Qin
Shannon Roberts
Jacqueline You
Rose Kelly
Peter Flockhart
Mina Michael
Shreya Pandit

CITY of BOSTON
Greater Boston Chamber of Commerce
MBA Massachusetts Bankers Association
Eastern Bank
100% TALENT
BOSTON UNIVERSITY
NSF
HRI

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

Software & Application Innovation Lab
MOC

Supported in part by NSF Awards #1430145, #1414119, #1718135, and #1739000; supported in part by the IARPA HECTOR program.

https://multiparty.org
fjansen@bu.edu