

Cloud-Enabled Big-Data Analytics: The Sky's the Limit

Managing Trust in the Cloud

Kinan Dak Albab

Rawane Issa

PhD Students, Boston University
Hariri Institute for Computing
Software Application & Innovation Lab
<https://multiparty.org/>

Trust and Computer Science

1. Public to trust correctness, security, and privacy of our solutions.
2. Data contributors to trust our algorithms and tools.
3. Policy makers to trust correctness and accountability of our algorithms and models.
4. Computer Scientists and researchers to trust the community and each other.
5. Business leaders, service providers, and resource providers to trust each other.

Can we compute the salary gap among genders and ethnicities without violating companies privacy concerns and revealing employee's records ?

During a global “cyber-attack” like the WannaCry ransomware attack, can we tell how far (in a network) we are from an infected PC without revealing if we are infected ourselves ?

Can we create a dating app which checks if two people are interested in each other but does not reveal to any party possible unreturned romantic interest ?

Can we conduct statistical analysis on medical data without revealing patient's privacy ?

YES we can.

Trust Management in the Cloud is Critical !

1. Sensitive / Valuable (“Big”) Data
2. Multiple computing parties / service providers
3. Multiple maintainers / managers of machines
4. Security of Data and Service

Who to trust?

Data Contributor:

1. Algorithm / Code
2. Service providers
3. Cloud providers
4. Code delivery mechanisms and CDNs
5. Computing stack

Service Provider:

1. Developers / Software
2. Data Contributor
3. Cloud Providers
4. Code delivery
5. Computing stack

Everybody has to trust
everybody else!

Cryptography is as much a **Social Science**
as it is a **Mathematical/Computer Science**.

Traditional “Solutions”

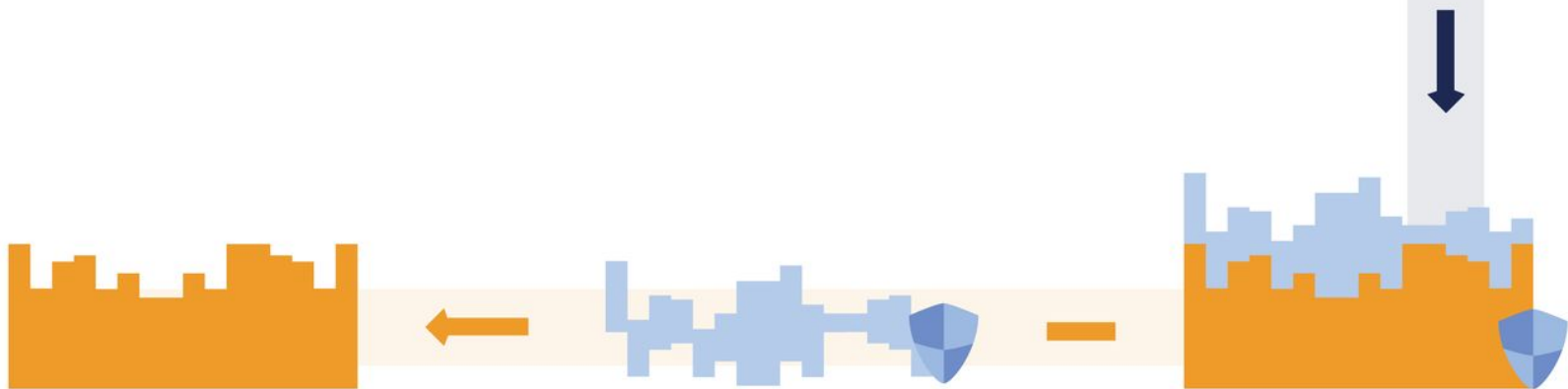
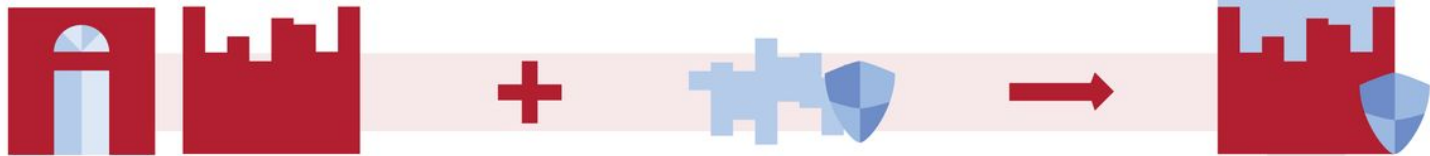
1. Trust blindly
2. Contract, regulation and policy
3. Anonymization

Secure Multi-Party Computation (MPC)

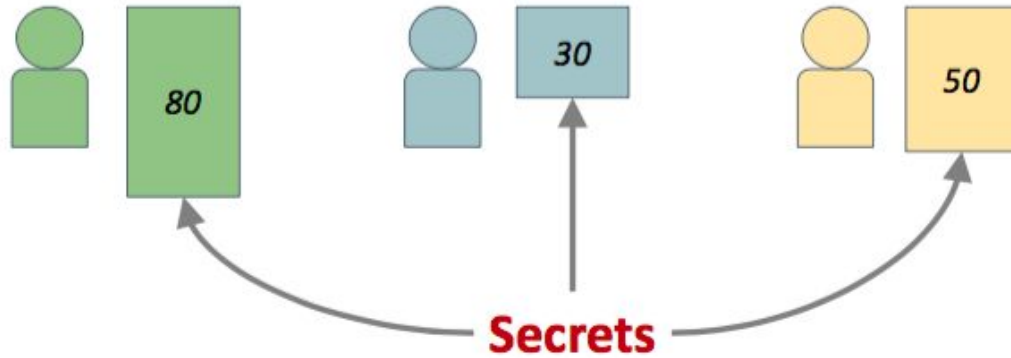
Sharing knowledge without sharing data*


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

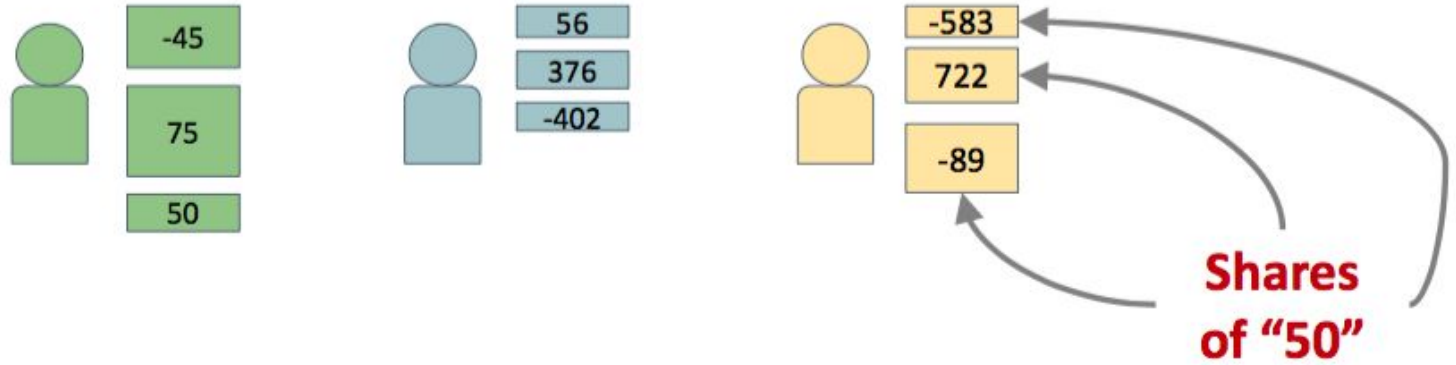
** under certain security assumptions*



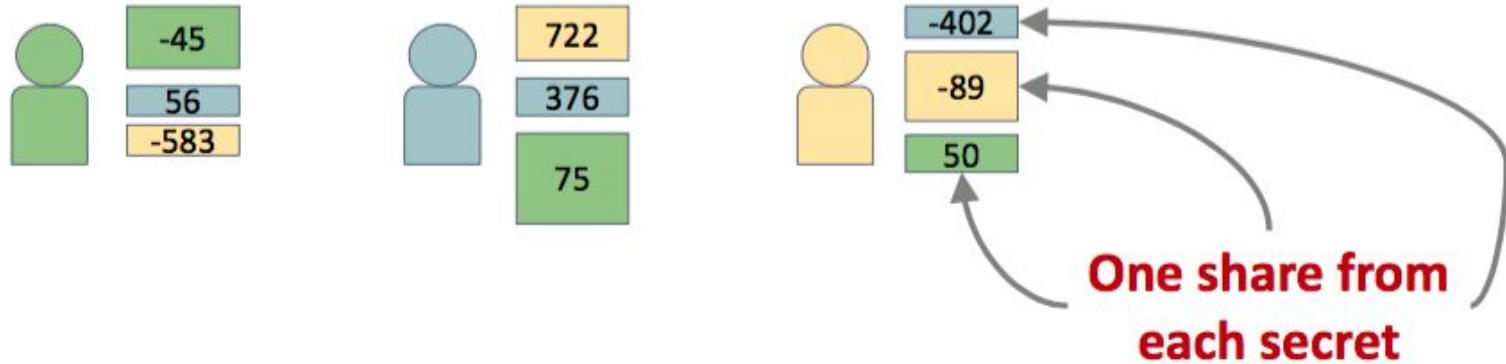
Sum all numbers without revealing them



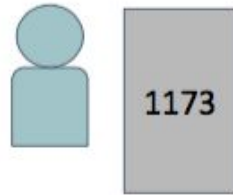
Compute secret shares



Send shares to corresponding parties



Sum received shares



Sum resulting three shares

-572

1173

-441

Sum = 160

Pay Equity

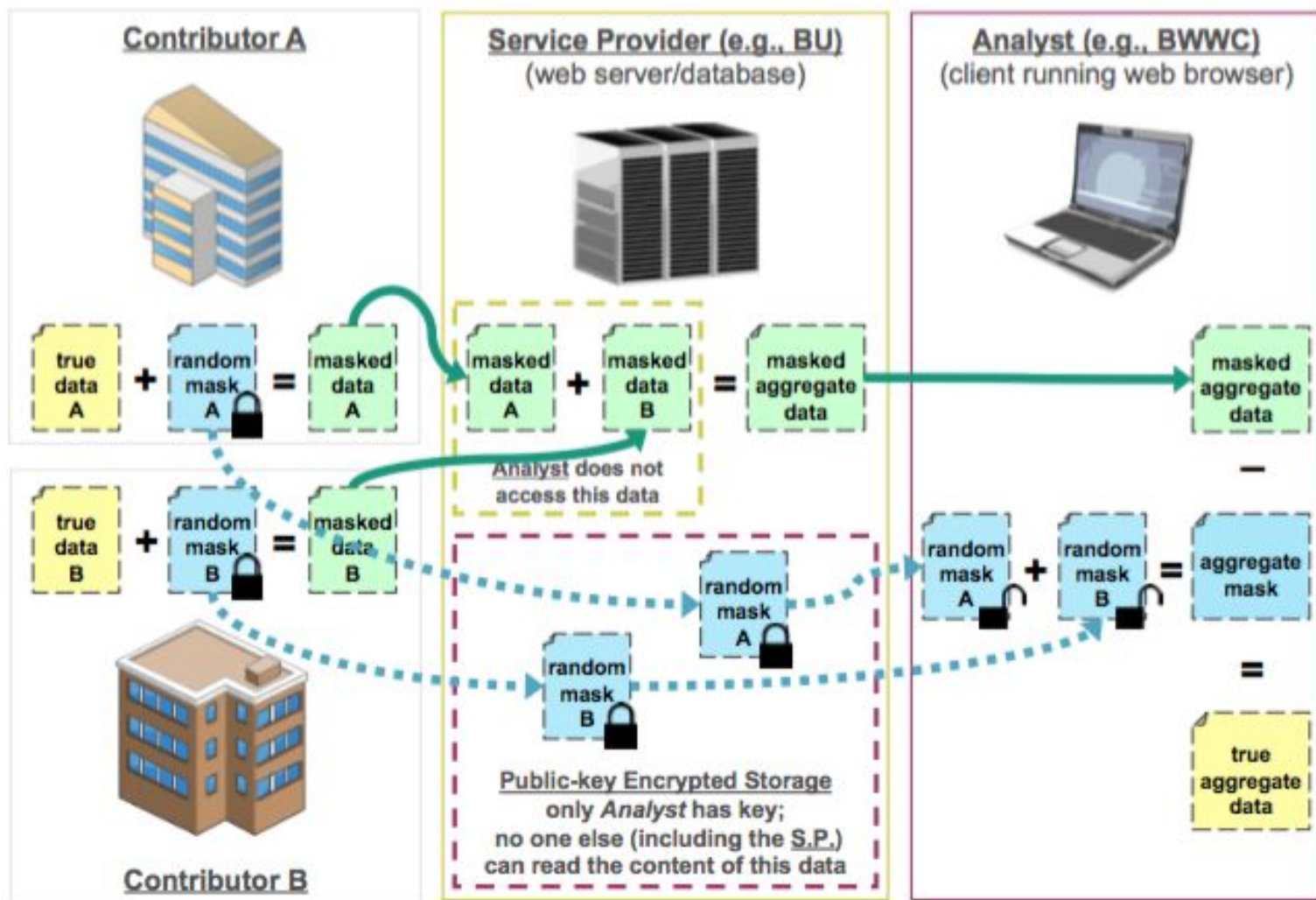
1. Use MPC to aggregate salary and employee information from many companies in the Boston area.
2. Analyse the aggregate to study pay equity between genders and ethnicities.












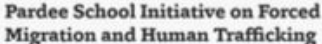
The Boston Globe

CITY of BOSTON





Current Applications

Partner(s)	Application(s)	Stage
 	<ul style="list-style-type: none"> Secure aggregation of tabular and multiple choice response data from multiple companies In upcoming deployment: correlations between multiple choice responses 	<ul style="list-style-type: none"> Deployed twice (5/2015, 6/2016) Actual result data published in report by BWWC (1/2017) Upcoming deployment (9/2017)
 	<ul style="list-style-type: none"> Secure aggregation and analysis of tabular data from multiple companies 	<ul style="list-style-type: none"> Implementation ready Deployment planned (early 2018)
 	<ul style="list-style-type: none"> Secure aggregation of tabular data and multiple choice response data from multiple banking organizations 	<ul style="list-style-type: none"> Implementation ready Awaiting deployment timeline
	<ul style="list-style-type: none"> Machine learning over sensitive consumer data subject to consumer-specified policies to enhance web services (e.g., route recommendation) 	<ul style="list-style-type: none"> Prototypes under development
 	<ul style="list-style-type: none"> Enhancement of mobile health intervention apps and data sharing tools with secure aggregate to add value for users and clinicians/researchers 	<ul style="list-style-type: none"> Application features under development Deployment planned (2018)
	<ul style="list-style-type: none"> Secure aggregation of tabular data from multiple organizations 	<ul style="list-style-type: none"> In discussions with stakeholders

Long Term Vision

Contributor

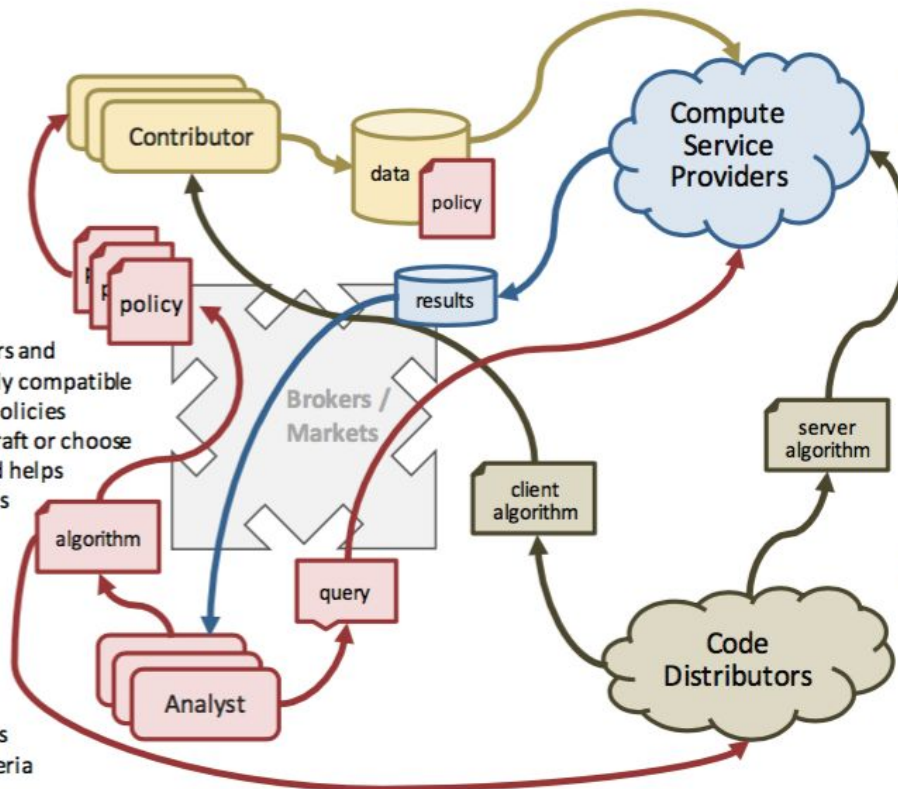
- Makes data available
- Sets/chooses policies

Broker/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

Collaborators and Research Group

“Researchers at Boston University, together with collaborators at several other institutions and organizations, are developing **open-source libraries, frameworks, and systems** that enable the implementation and deployment of applications that employ secure multi-party computation in accessible and scalable ways.”

<https://multiparty.org/>

Please **contact us** if you would like to learn more or are interested in collaborating.

Thank you!

babman@bu.edu ra1issa@bu.edu