Secure Multiparty Computing for Business Data

A new Big Data Paradigm to collect Economic Data

Cavan Capps
Big Data Lead
U.S. Census Bureau
September 13, 2016

Data is for decision making

It's value is on based on timeliness and relevancy

Economic data often lags economic activity or is only collected nationally. Much of the economy is MSA based.

Confidentiality becomes more critical

No longer small samples

So much data to match to…

- New Techniques: Formal Privacy & Differential Privacy
  - Apple and Google using Differential Privacy to collect data
- Secure Multi-party Computing

Data Holes During the Last Recession

- One Example:
  - Companies were facing bankruptcy (GM)
  - No supply-chain data
  - No sense of what other businesses or jobs in other states would be affected

Need for Modernized Data Collection

- Every Business Transaction is already Electronic
- Obvious solution: Passive collection of Business electronic transactions – rather than only paper questionnaires

Trust Challenges

Timely Data is more valuable

- There are increased security issues with timely data
- There are increased “hacks” from “Advanced Persistent Threat” attackers
- Global Competition heightens fears of broken confidentiality.

Can we have Confidentiality & Security with "Near Real Time" access to data

- Real Time Data Streams/Data Dumps
  - Access to unencrypted transaction data may provide exposure to attacks from:
    - Cyber-security hacks
    - Confidentiality threats.

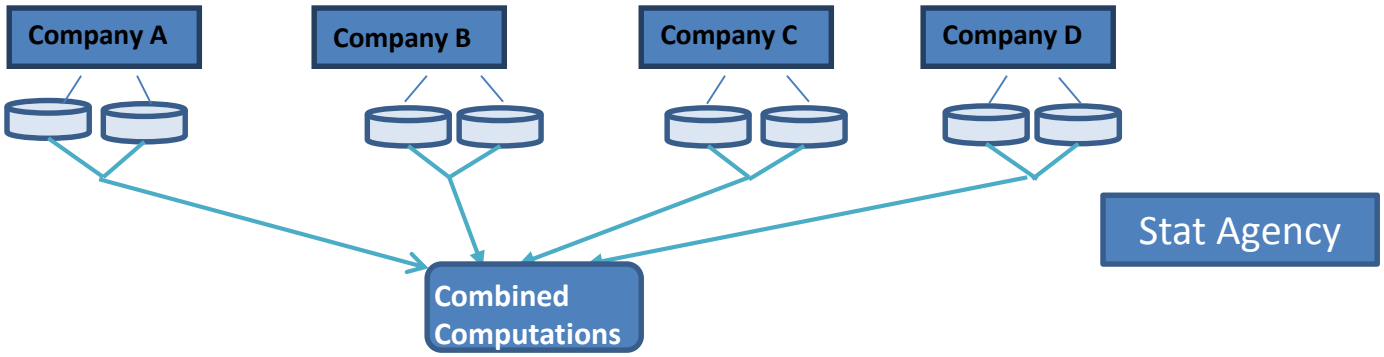
A Proposal for Secure Multi-party Computing

Companies may not trust government to process real time data. Real time data has more critical proprietary information in it.

Stat Agency

Currently companies process proprietary data in separate data systems. Typically the data is in databases designed to process business transactions, designed to find and update the correct record out of a large set of business records.

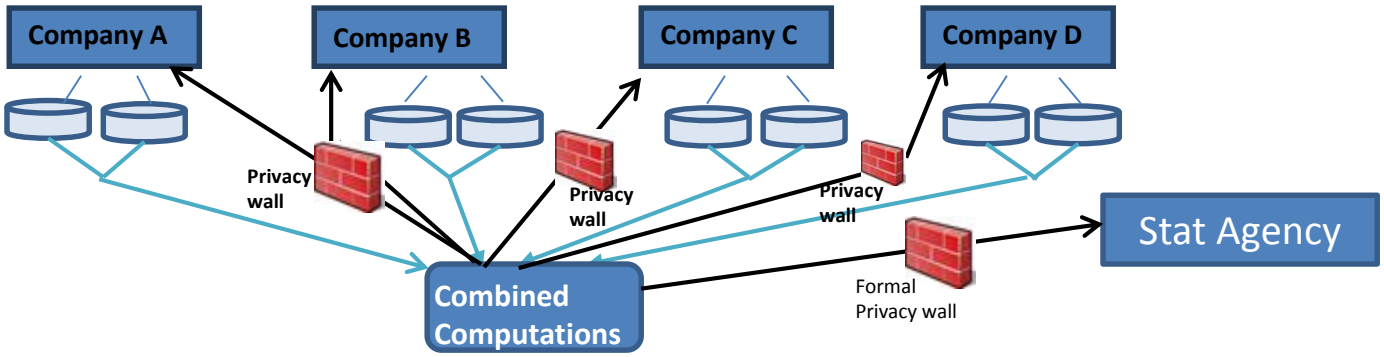
Secure Multi-party Computing



In a secure multi-party computing system all ***computations are done on encrypted data***. Keys are set up so different actors can access different views of the data.

Each party can see and analyze their data but can not see anyone else's data on the network.

Secure Multi-party Computing

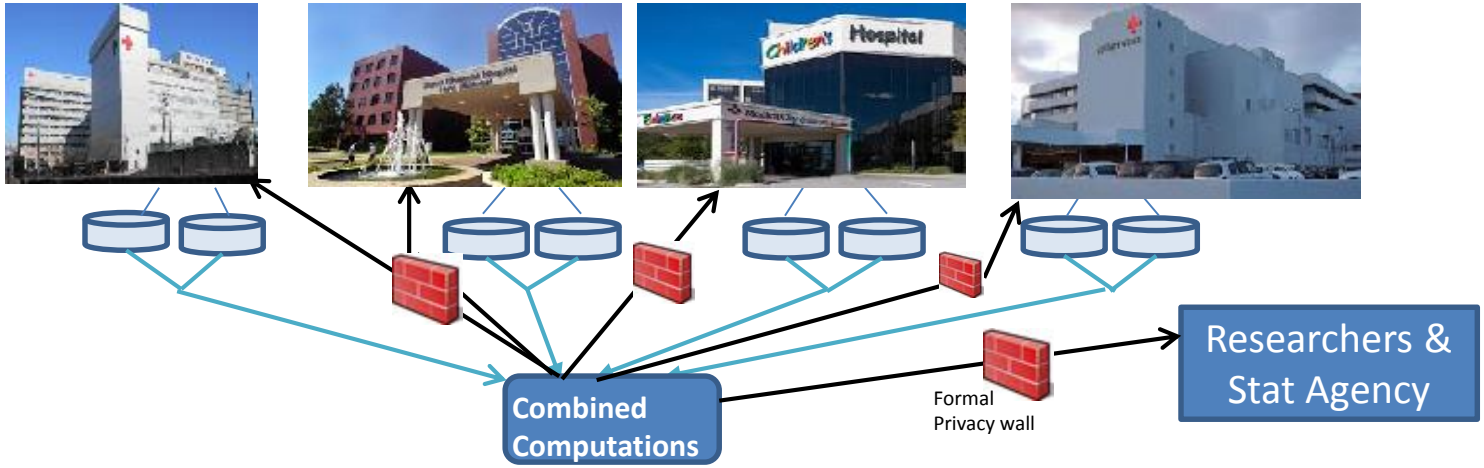


Certain types of analytical results are available for the entire industry by using distributed processing across the network of industry analytical databases.

All results from Industry-wide(multi-party) analysis are formally privatized before being reported. Neither the participants or the Statistical Agency is able to view results without going through a “Formal Privacy Filter”.

Questions can be asked of the network of computers without seeing any individual record.

Secure Multi-party Computing

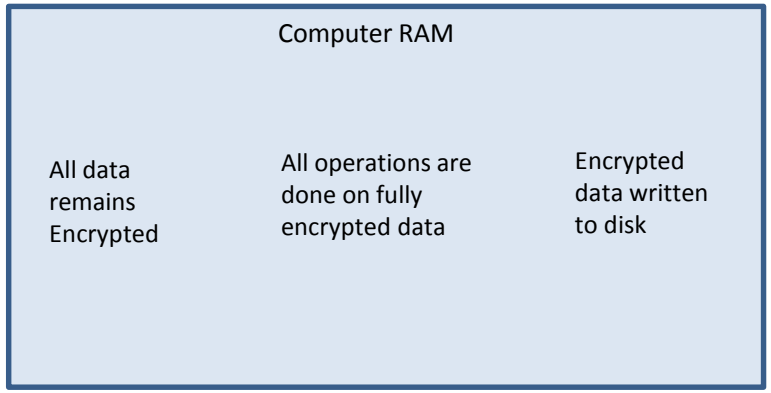


Similarly, researchers might be able to do analysis across a set of businesses and other company records while keeping data strictly private. This would by pass the need to move all records to a complex central database which becomes a central hacking target. Decentralized encrypted records should also reduce “Big Brother” concerns.

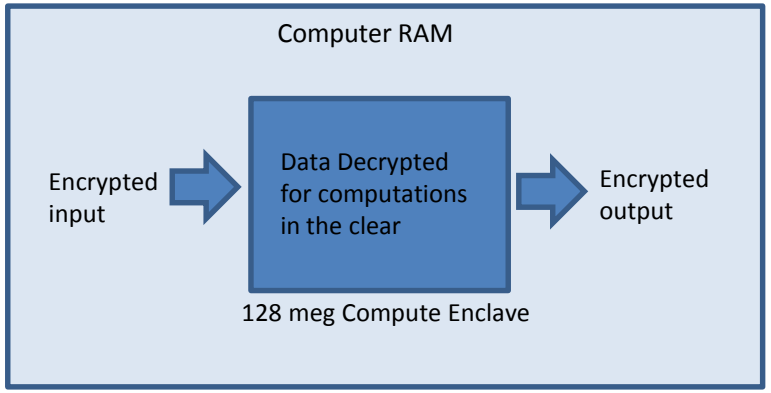
The **enforced privacy will also simplify access to data** for advanced research.

Two SMC Options

Homomorphic Encryption



SGX Option



Types of potential Analyses

- Imputation
- Tabulation
- Distributed Regression
- Machine Learning(Random Forests, Support Vector Machines, Text Mining…)
- Distributed Linkages using Tokens

Trust may require Open Source

- Software solutions must be transparent and be open to review by Industry

Free-Rider Challenges

- The Free Rider Problem occurs when those who benefit from resources, goods, or services do not pay for them, which results in an under-provision of those goods or services
- Does industry face joint problems where collaboration across companies make sense such as:
  - In the shipping industry where there is are “Alliances in the shipping area for large containers, movement to regional distribution centers, bottlenecks in shipping infrastructure, early identification of shipping structure bottlenecks as technology changes.” Mckinsey & company

Questions/Suggestions??

- Cavan.Paul.Capps@census.gov

More information:

--- Altman M, Capps C, Prevost R. Location Confidentiality and Official Surveys. Social Science Research Network [Internet]. 2016.

http://projects.informatics.mit.edu/bigdataworkshops/publications/location-confidentiality-and-official-surveys

--- Workshop presentations:

http://projects.informatics.mit.edu/bigdataworkshops/book/export/html/558