Secure MPC as a Tool for Sensitive Data Analysis

Samuel Havron ⟨havron@virginia.edu⟩

February 15, 2016

1 Introduction

A *Secure multi-party computation* (MPC) is a protocol which allows for two or more parties to perform a function (typically a mathematical computation) on sensitive input data provided by each party, without revealing anything about the inputs. At an abstract level, many current implementations of MPC work by executing instructions *data-oblivious* manner, where the control flow of the program is independent of the inputs provided by each party. MPCs have practical uses for many organizations that have sensitive data sets which they cannot share, but could otherwise learn a lot by computing statistical functions on their joint data. With a computer program that implements MPC, parties can safely share their sensitive data in such a manner that only the output of the program is revealed to them, with all sensitive input and intermediary data being hidden. For instance, two competitor companies may want to compute statistical functions on joint sales data and publish the results for internal company research and marketing, without allowing either competitor to uncover sensitive sales data about the other. In such situations it is important that neither company know the sales data inputs of their competitor, as they may learn secrets or be able to manipulate the data in their favor.

- Capabilities of MPC (including overview of previous work, Samee's implementation, efficiency/speed comparison with non-secure MPC programs)
- Value of MPC to social scientists, researchers (with motivating example(s)) as a tool for sensitive data analysis

2 Methods

- Description of OblivCs efficiency/speed in comparison to other existing MPC implementations
- Overview of OblivC as a language, relationship to C (appeal to researchers with some C/basic programming experience)

- Introduce implementation of linear regression analysis program, use code snippets and reference repo/site tutorial

3 Results

- Discuss efficiency/speed/scalability of aforementioned linear regression analysis program, provide EC2 testing data (collect more formally prior to including in final paper)
- Discuss results of other motivating OblivC programs, compare to alternative MPC results

The scalability and speed of OblivC as a tool for implementing MPC programs was tested using *Elastic Compute Cloud* (EC2) nodes of varying tiers from *Amazon Web Services* (AWS) of Amazon.com®. Data for computation was mined from New York public health data (include link), as well as synthetic data points generated through a Python program (footnote for file location) for performance results.

Table 1: Performance of MPC linear regression analysis

Input Size	EC2 Node Tier	Time to Compute
1,000	GP:t2.micro	11.755s
5,000	GP:t2.micro	57.997s
10,000	GP:t2.micro	115.089s
1,000	GP:m4.large	11.262s
5,000	GP:m4.large	55.233s
10,000	GP:m4.large	110.049s
1,000	CO:c4.xlarge	5.783s
5,000	CO:c4.xlarge	27.928s
10,000	CO:c4.xlarge	57.433s

Data collected represents average compute time *(5 iterations per test)*.

- Include table with times from non secure MPC computation for comparison.

4 Conclusion

- Overview of secure MPC, utility/extensibility of OblivC programming, real-world usage for scientists/researchers