

CPSC 490 Prospectus

Will Childs-Klein

Introduction

Parallel data processing methods have allowed people of all professions to analyze data at an unprecedented scale, yielding key insights in myriad industries and academic science. Contemporaneously, and not entirely independently, data privacy has become an important and immediate issue in our society. Organizations in both the public and private sectors constantly collect or access data on their users or constituents, often without explicit consent by the user. As computing continues to shift into “the cloud” from users’ local systems, user data is relinquished to and becomes centralized in the small number of prevalent cloud service providers. In distributed parallel data processing, where clusters of multiple physical or virtual machines execute jobs in parallel, this introduces a high potential for privacy and confidentiality of users’s data and the results of their computation. Even assuming an entirely benign cluster provider, the contemporary model of centralized clusters provides a focus point of attack for malicious hackers. In this project, we will focus on the MapReduce paradigm

Problem Statement

So, how can we protect the privacy of users’ computations while still affording the user all (or at least most) of the major features of modern distributed data processing architectures? We would need a system that:

1. is decentralized (peer-to-peer) in architecture with granular data access control, to prevent a malicious peer from being able to see too great a portion of the user’s data.¹
2. provides the user anonymity² from other peers on the network
3. maintains compatibility with existing MapReduce implementations without forcing the user to re-write large portions of their code base.

¹this may be susceptible to sybil attacks

²or at least pseudonymity

Deliverables

1. code for simulations/proof of concept
2. code documentation
3. formal document containing description of project, implementation details & design decisions, and analysis of experimental results.